Monotone Likelihood, Powerfunction Diagrams and Selection

By Erik Torgersen
University of Oslo

ABSTRACT

This is a study of monotone likelihood (M-L) experiments i.e. experiments which possess monotone likelihood in some statistics. The main tools are the concave functions, here called Neyman-Pearson (N-P) functions, which describe the relationships between level of significance and minimum power for maximin tests.

If the parameter set contains two points, i.e. in the case of dichotomies, these functions describe the experiment up to equivalence. Structures for sets of dichotomies are often very simply expressed in terms of Neyman-Pearson functions. It turns out that several of these structures of dichotomies extend naturally to M-L experiments. Thus, for example, the set of types of M-L experiments is order complete for the pairwise ordering and it is compact for the weak experiment topology.

Types of M-L experiments are determined by families of N-P functions satisfying obvious consistency requirements. These requirements may be expressed as a semigroup property of N-P functions for functional composition. A closely related representation is in terms of powerfunctions of most powerful tests.

Using these representations we consider comparison, exact or approximate, of one M-L experiment w.r.t. another experiment. Generalizing results in Lehmann (1988) we show that comparison for monotone decision problems reduces to pairwise comparison i.e to dichotomies. By general comparison principles, for given classes of loss functions, this extends to products and mixtures of M-L experiments. In particular we obtain comparison results for the n-sample case.

Other interesting characterizations and representations are in terms of supports of standard measure, in terms of the between property for statistical distance and, in the differentiable case, in terms of local comparison. In the last case M-L experiments are characterized by families of functions, here called slope functions, providing slopes of powerfunctions of most powerful tests.

Besides unavoidable continuity/differentiability conditions there are no consistency requiremention for these families of slope functions. The type of a differentiable M-L experiment be recovered from the slope functions by solving first order differential equations.

The results are used to explore how information is affected by selection. If the random variable X constituting our original (ideal) experiment are observable only when a given

1. INTRODUCTION. MONOTONE LIKELIHOOD AND MONOTONE DECISION PROBLEMS. ORDERING OF EXPONENTIAL MODELS.

In a most interesting paper Bayarri and DeGroot (1987) consider how information is influenced by selection. Most of the models discussed by them were monotone likelihood (M-L) experiments. As this property is preserved by selection the general principles for comparing such experiments apply. In particular the important criterion of Lehmann (1988) for comparison for monotone decision problems applies in the non atomic case. A general criterion for the overall comparison of such experiments which effectively utilizes the M-L property is still not available.

If the models are exponential (Darmois-Koopman) then overall comparison may by Janssen (1988) be decided by a convolution criterion. This important criterion is described later in this section.

We shall here mostly discuss comparison where at least one of the experiments under comparison has the monotone likelihood property. We shall however also make a few comments on general comparison, and in particular on comparison for a given loss-function.

The point of departure of this paper is to consider the information stored in the set of powerfunctions of one sided tests.

In the case of dichotomies this set completely characterizes the type (i.e. the equivalence class) of the experiment. Moreover it provides a simple canonical representation having interesting properties. As any dichotomy may be considered as a M-L experiment, we begin in section 2 by briefly describing some basic properties of dichotomies.

Dichotomies may be studied in terms of their *Neyman-Pearson (N-P)* functions. These are the functions which relate level of significance to maximin power. They abound in mathematical statistics and are, up to trivial modifications, nothing but the Lorentz transformations (curves) from econometry or the total time on test (TTT) transform of reliability theory. Both aspects are statistically interesting. In this paper we shall in particular see how basic tools from reliability theory may be explored to obtain general information on statistical models.

In the case of differentiable experiments we may instead consider functions which associate maximum slopes of powerfunctions of tests with given levels of significance. These functions, here called *slope functions*, determine, as explained in Torgersen (1985), all local information. In particular they determine Fisher information. They enter naturally into the discussion here and we shall therefore provide a brief exposition in section 3.

After these preparations we turn to general M-L experiments in section 4.

We begin here by observing that the M-L property amounts to the requirement that the likelihoods are totally ordered for the natural ordering. This implies readily that the M-L property is a triplewise property, i.e. that it suffices to consider subsets of the parameter set having three points.

Combining this with the standard method for obtaining minimum Bayes risk we conclude that if $\mathcal{E}_i$ i=1,...,r, is at least as informative as $\mathcal{F}_i$ for a given class of lossfunctions then the product $\mathcal{E}_1 \times \cdots \times \mathcal{E}_r$ is at least as informative as $\mathcal{F}_1 \times \cdots \times \mathcal{F}_r$ for the same class of loss functions.

A similar statement holds for mixtures of experiments.

As pointed out in Lehmann (1988) the overall comparison may be afflicted by particular decision problems which are not of interest in a given situation. LeCam (1964) provides criteria which may throw light on what kind of decision problems are responsible for the noncomparability og given experiments.

The problem of finding methods for comparing experiments w.r.t. given decision problems of particular interest is a challenging one. In addition to the references in Lehmann (1988) we would also like to point out that: comparison for k-decision problems is treated in Blackwell (1953) and Torgersen (1970), comparison for estimation of linear parametric functions were discussed in Stepniak and Torgersen (1981), in Stepniak, Wang and Wu (1984), in Torgersen (1984 and 1990) and in Swensen (1980).

The theory of LeCam (1964 and 1986) provides criteria for asymptotic local comparison of risk functions for given decision problems. Fixed sample size local comparison is discussed in Torgersen (1972a-b and 1985).

In the last part of section 6 and in section 7 we apply the obtained criteria, as well as other ones, to experiments obtained by selection. We shall here mainly be concerned with orderings of experiments related to one of the following modes of comparison:

Over all comparison = comparison on all of Θ for all decision problems.

Local comparison = comparison within small neighbourhoods of any given parameter point and for all decision problems. (If Θ is one dimensional then it suffices to consider testing problems).

m-wise comparison = comparison for all m-point sub parameter sets and for all decision problems.

Pairwise comparison = 2-wise comparison = comparison for all two point subparameter sets and for all decision problems (it suffices to consider testing problems).

Other orderings which will be considered are the natural orderings defined by the Hellinger transform, by affinities and by Fisher information. These orderings are related as in the figure below, where arrows point in the direction of decreasing strength:

By the previous sections comparison for monotone decision problems may be expressed in terms of N-P functions or in terms of slope functions. In the first case the problem is reduced to pairwise comparison and in the second to local comparison. Thus we shall take a look on how N-P functions and slope functions are altered by selection. In the exponential case, we do the same thing for the Hellinger transform.

We shall in section 7 restrict our attention to selection on subsets of the likelihood space which are either intervals or complements of intervals. It turns out, in both cases, that whether information is reduced or increased is related to whether the failure rates of laws of (differentiated) log likelihoods are increasing or decreasing. Among the results we mention that selection on an interval bounded away from 0 and infinity in the likelihood space never increases strictly the pairwise information.

In section 8 we return to the general situation and consider conditions ensuring that a M-L experiment approximately majorizes another experiment for monotone decision problem.

In the case of dichotomies this is expressed in terms of the deficiency introduced by LeCam (1964). This deficiency has, as shown in Torgersen (1970), natural geometrical interpretations. In particular the deficiency distance/$\sqrt{2}$ is the Paul-Levy diagonal distance between these functions considered as distribution functions.

The main results in this section extend this by showing that a M-L experiment is ϵ - deficient w.r.t. another experiment for monotone decision problems if and only if this is so pairwise.

At this point we should make precise what is here meant by monotone likelihood ratio, by monotone decision problems and by monotone decision procedures.

Consider an experiment $\mathcal{E} = (\chi, \mathcal{A}; P_\theta; \theta \in \Theta)$ along with a real valued statistic Z on the sample space $(\chi, \mathcal{A})$ of $\mathcal{E}$. Assuming that Θ is a set of real numbers we shall say that $\mathcal{E}$ *has monotonically increasing (decreasing) likelihood ratio in Z* if there to each pair $(\theta_1, \theta_2) \in \Theta \times \Theta$ such that $\theta_2 > \theta_1$ corresponds a monotonically increasing (decreasing) function $\varphi_{\theta_2,\theta_1}$ on R such that $\varphi_{\theta_2,\theta_1}(Z)$ is a P_{θ_2} maximal version of $dP_{\theta_2}/dP_{\theta_1}$.

Note that maximality implies that $\varphi_{\theta_2,\theta_1}(Z) = \infty \quad a.s. \quad P_{\theta_2}$ on any P_{θ_1} null set N.

Clearly $\mathcal{E}$ has monotonically increasing likelihood ratio in Z if and only if $\mathcal{E}$ has monotonically decreasing likelihood ratio in -Z. Thus we may in many problems restrict our attention to monotonically increasing likelihood ratios.

Consider now the particular case of a totally informative experiment i.e. an experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ such that P_{θ_1} and P_{θ_2} are mutually singular when $\theta_1 \neq \theta_2$. If the sample space χ permits a measurable patitioning $A_\theta : \theta \in \theta$ such that $P_\theta(A_\theta) \equiv 1$ then $\mathcal{E}$ has monotonically increasing likelihood ratio in Z where $Z(x) = \theta$ when $x \in A_\theta$ and $\theta \in \Theta$. Although this construction is feasible whenever Θ is countable, and thus for any restriction to a countable subparameter set, there may not be any statistic Z such that $\mathcal{E}$ has

If conditions (i) and (iii) hold while (ii) is violated with τ being monotonically decreasing then we may replace Θ with $\tilde{\Theta} = -\Theta$ and L with $\tilde{L}$ defined by $\tilde{L}_{\tilde{\theta}}(t) = L_{-\tilde{\theta}}(t)$ when $\tilde{\theta} \in \tilde{\Theta}$ and $t \in T$. Then $\tilde{L}$ is monotone with τ being replaced by $\tilde{\tau}$ given by $\tilde{\tau}(\tilde{\theta}) = \tau(-\tilde{\theta})$; $\tilde{\theta} \in \tilde{\Theta}$.

If L is monotone and the decision space T is finite, say $T = \{t_1, ..., t_k\}$ with $t_1 < \cdots < t_k$, then we may decompose Θ as $\Theta = \Theta_1 \cup \cdot\cdot \cup \Theta_k$ where $\Theta_i = \{\theta : \tau(\theta) = t_i\}$. Then $\Theta_1, ..., \Theta_k$ are disjoint intervals in Θ and $\theta_1 < \cdots < \theta_k$ whenever $\theta_i \in \Theta_i; i = 1, .., k$. If, in addition, Θ is finite then there are parameter points $\theta_1 < \theta_2 < \cdots < \theta_{k-1}$ so that:

$$\Theta_1 = \{\theta : \theta \in \Theta \quad and \quad \theta \leqq \theta_1\}, \quad \Theta_2 = \{\theta : \theta \in \Theta \quad and \quad \theta_1 < \theta \leqq \theta_2\}, \cdots,$$

$$\Theta_{k-1} = \{\theta : \theta \in \Theta \quad and \quad \theta_{k-2} < \theta \leqq \theta_{k-1}\} \quad and \quad \Theta_k = \{\theta : \theta \in \Theta \quad and \quad \theta > \theta_{k-1}\}$$

A general monotone decision problem may be approximated by a finite monotone decision problem by the following device. Let (T, L) be monotone as described above and consider any non empty subset T_0 of T and any non empty subset Θ_0 of Θ. Then there are subsets T_1 of T and Θ_1 of Θ such that $T_0 \subseteqq T_1, \Theta_0 \subseteqq \Theta_1$ and $L|\Theta_1 \times T_1$ is monotone. In fact we may put $T_1 = T_0 \cup \tau[\Theta_0]$ and $\Theta_1 = \Theta_0 \cup \{\theta_t : t \in T_1 - \tau[\Theta_0]\}$ where $\theta = \theta_t$, for each t, is a solution of the equation $\tau(\theta) = t$. Then $\tau[\Theta_1] = T_1$ so that $L|\Theta_1 \times T_1$ is monotone.

Note that the sets Θ_1 and T_1 are both finite when the sets Θ_0 and T_0 are finite.

If the experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ has monotonically increasing likelihood ratio in a statistic Z and if the loss function is monotone then, by Karlin and Rubin (1956), we may restrict attention to decision rules δ which are *monotone* in the sense that $\delta([t, \infty[|Z = z) = 1$ whenever $\delta([t, \infty[|Z = z') > 0$ for some $z' < z$.

If δ is non randomized this amounts to the condition that $\delta(z)$ is monotonically increasing in z. For randomized procedures the above definition of monotonicity amounts to a stronger requirement than the requirement that $\delta(\cdot|z)$ increases monotonically in z for the stochastic ordering of distributions.

So far no simple general criteria for the overall comparison of M-L experiments appear to be available. In two important particular cases convolution criteria apply.

Firstly if $\mathcal{E}$ is a strongly unimodal translation experiment then the convolution criterion of Boll (1955) for comparability of translation experiments applies. By Torgersen (1972) this extends to approximate comparison.

Secondly if $\mathcal{E}$ is an exponential (Darmois-Koopmans) family of any finite dimension then the convolution criterion of Janssen (1988) is available. (This result is also close to the surface in Ehm and Müller (1983), although the emphasize in that paper is on asymptotic comparison).

One dimensional exponential experiments are M-L experiments and, furthermore, exponentiallity is preserved by selection. Thus the latter criterion is very relevant here. As it may not be so well known we shall provide the following version of it.

for the distribution $P_{\theta_0} \times M$. Then $P_{\theta_0} \times M = D \times Q_{\theta_0}$ and $P_{\theta_0} = DQ_{\theta_0}$. It follows that D maps densities dP_θ/dP_{θ_0} into versions $z \to \int [dP_\theta/dP_{\theta_0}]_y D(dy|z)$ of dQ_θ/dQ_{θ_0}. In other words $\int \frac{a(\theta)}{a(\theta_0)} e^{(\theta-\theta_0,y)} D(dy|z) = \frac{b(\theta)}{b(\theta_0)} e^{(\theta-\theta_0,z)}$ for Q_{θ_0} almost all $z \in R^k$. Let D_z be the distribution of $Y - z$ when Y is distributed according to $D(\cdot|z)$. Then $\int e^{(\theta-\theta_0,x)} D_z(dx) = \int e^{(\theta-\theta_0,y-z)} D(dy|z) = \frac{a(\theta_0)b(\theta)}{a(\theta)b(\theta_0)}$ for Q_{θ_0} almost all z.

By continuity and separability we may arrange the exceptional set of points z such that it does not depend on θ as long as θ belongs to an open ball around θ_0. Using analytic continuation we find that the characteristic function of D_z, for z not belonging to this exceptional set, is independent of z and thus that $D_z = G$ where the distribution function G does not depend on z.

For θ belonging to the above mentioned ball around θ_0 this implies $\int e^{(\theta-\theta_0,x)} G(dx) = \frac{a(\theta_0)b(\theta)}{a(\theta)b(\theta_0)}$ i.e. that $\frac{a(\theta_0)}{a(\theta)} = \frac{b(\theta_0)}{b(\theta)} \int e^{(\theta-\theta_0,x)} G(dx)$. The last equality may also be written

$$E_{\theta_0} e^{(\theta-\theta_0,Y)} = E_{\theta_0} e^{(\theta-\theta_0,Z)} \int e^{(\theta-\theta_0,x)} G(dx).$$

Hence

$$E_{\theta_0} e^{i(t,Y)} = E_{\theta_0} e^{i(t,Z)} \int e^{i(t,x)} G(dx); \quad t \in R^k$$

and thus:

$$\mathcal{L}_{\theta_0}(Y) = \mathcal{L}_{\theta_0}(Z) * G.$$

□

Among the notations which will be used are:

$N(\xi, \sigma) =$ the univariate normal distribution with mean ξ and standard deviation σ.

$R(0,1) =$ the uniform (rectangular) distribution on [0,1].

$\mathcal{L}(X) =$ distribution (law) of X.

DFR= decreasing failure rate.

IFR= increasing failure rate.

$d\nu/d\mu =$ the Radon-Nikodym derivative of the μ absolutely continuous part of ν w.r.t μ.

$a \wedge b =$ minimum $\{a, b\}$

$a \vee b =$ maximum $\{a, b\}$

$\bigwedge_t a_t = \inf_t a_t$

2. THE CASE OF DICHOTOMIES. NEYMAN-PEARSON (N-P) FUNCTIONS.

Any dichotomy $\mathcal{D} = (P_0, P_1) = (P_\theta; \theta \in \{0,1\})$ is a M-L experiment. It is therefor natural to begin our study of M-L experiments by reviewing properties of dichotomies. Convenient tools in this case are:

(i) The relationship between level of significance and maximum power for testing, say, "$\theta = 0$" against "$\theta = 1$".

(ii) The relationship between prior distribution and minimum Bayes risk for testing "$\theta = 0$" against "$\theta = 1$" with 0-1 loss.

(iii) Variations of standard measures and Blackwell measures.

(iv) The Hellinger transform.

The relationship (i) is given by functions which in one form or another, appear to play important roles at the most diverse occasions, not all of them in statistics. Although not widely recoqnized, even among statisticians, their genesis may be regarded as rooted in the Neyman-Pearson lemma. They deserve a name expressing this and we shall here say that a function is a *Neyman-Pearson function (N-P function)* if it is a continuous concave function from the unit interval [0,1] to itself which leaves 1 fixed. Of course concavity ensures continuity on the open interval]0,1[and if, in addition, it is assumed that 1 is a fixed point then it is automatically continuous on]0,1]. Thus a function β from the unit interval to itself is a N-P function if and only if it is concave, $\beta(0+) = \beta(0)$ and $\beta(1) = 1$.

In statistics N-P functions arise in testing theory in many situations which are not directly related to the Neyman-Pearson lemma. Thus e.g. the maximin level α power defines a N-P function β of α provided we ensure that $\beta(0+) = \beta(0)$.[If the weak compactness lemma holds then this is automatic. In general we may just define $\beta(0)$ as $\beta(0+)$.]

More generally we may consider maximin level α power for test functions belonging to a given convex class of test functions containing the constants in [0,1].

In particular if $\mathcal{D} = (P_0, P_1)$ is a dichotomy then *the N-P function* of $\mathcal{D}$ is the function $\beta(\cdot|\mathcal{D})$ which to each $\alpha \in [0,1]$ assigns the power $\beta(\alpha|\mathcal{D})$ of the most powerful level α test for testing "$\theta = 0$" against "$\theta = 1$". When convenient this function may also be denoted as $\beta(\cdot|P_0, P_1)$.

Example 2.1 (Double dichotomies and triangular N-P functions).

If β is a N-P function then $\alpha \leqq \beta(\alpha) \leqq 1$ for all $\alpha \in [0,1]$. The lefthand side corresponds to the N-P function of a totally non informative dichotomy (P, P) while the right hand side corresponds to a totally informative dichotomy (P_0, P_1) with P_0 and P_1 being mutually singular.

An interesting family of N-P functions (which include the above mentioned) are the triangular ones. These are the N-P functions of the double dichotomies. Thus the N-P function of the double dichotomy $((1-p, p), (1-q, q))$ with $p \leqq q$ is the upper boundary of the we may write $\underline{\mathcal{D}} = \inf_i \mathcal{D}_i$ or

$$\underline{\mathcal{D}} = \bigwedge_i \mathcal{D}_i.$$

It follows that the collection of dichotomies is order complete for the informational ordering. Note however that the sup operation expressed for N-P functions is not the pointwise supremum. It corresponds of course to the supremum operation on N-P functions for the informational ordering. The least upper bound of a family $(\mathcal{D}_i : i \in I)$ of dichotomies may be denoted as $\sup_i \mathcal{D}_i$ or as $\vee_i \mathcal{D}_i$.

Monotone likelihood experiments are very naturally represented as families of N-P functions. These families are characterized by being closed for the "natural" functional compositions. In general if $\mathcal{D}_1$ and $\mathcal{D}_2$ are dichotomies having, respectively N-P functions β_1 and β_2 then the composed function $\beta_1(\beta_2) = \beta_1 \circ \beta_2$ is also a N-P function. If $\mathcal{D}$ is a dichotomy having $\beta_1(\beta_2)$ as its N-P function then $\mathcal{D}$ is at most as informative as the product dichotomy $\mathcal{D}_1 \times \mathcal{D}_2$. Indeed if γ is the N-P function of $\mathcal{D}_1 \times \mathcal{D}_2$ then, for any $\alpha \in [0,1], \gamma(\alpha) = \sup\{\int \beta_1(\alpha(x))\beta_2(dx) : \int_0^1 \alpha(x)dx = \alpha\} \geqq$ (by Jensen's inequality) $\sup\{\beta_1(\int \alpha(x)\beta_2(dx)) : \int_0^1 \alpha(x)dx = \alpha\} = \beta_1(\beta_2(\alpha))$.

As mentioned above any N-P function arises from a dichotomy. In fact a N-P function β is also a cumulative distribution function of a probability distribution on [0,1] which is absolutely continuous on]0,1]. In fact it may be checked that β is the N-P function of the pair $(R(0,1),\beta)$ where R(0,1) denotes the rectangular distribution on (0,1).

The N-P function of a dichotomy $\mathcal{D} = (P_0, P_1)$ is usually found by first finding a real valued sufficient statistic $X, e.g.X = dP_1/d(P_0 + P_1)$, such that $F_1 = \mathcal{L}(X|P_1)$ has a monotonically increasing density w.r.t. $F_0 = \mathcal{L}(X|P_0)$. By the Neyman Pearson lemma $\beta(\alpha|P_0, P_1) = 1 - F_1(F_0^{-1}(1-\alpha))$ for any $\alpha \in]0,1[$ such that $F_0(F_0^{-1}(1-\alpha)) = 1-\alpha$. In general this formula holds for any $\alpha \in]0,1[$ provided we permit a random mass in $F_0^{-1}(1-\alpha)$ distributed uniformly on $[0, m(\alpha)]$ where $m(\alpha)$ is the F_0 mass in $F_0^{-1}(1-\alpha)$.

All dichotomies having the same N-P function β are statistically equivalent with the dichotomy $(R(0,1),\beta)$. Using the terminology of LeCam (1986) we may express this by saying that $\beta(|\mathcal{D})$ defines the type of the dichotomy $\mathcal{D}$. In fact if $\hat{\alpha}$ is the observed significance level for testing "$\theta = 0$" against "$\theta = 1$" in $\mathcal{D} = (P_1, P_0)$ then $\mathcal{L}(\hat{\alpha}|P_0) = R(0,1)$ and $\mathcal{L}(\hat{\alpha}|P_1) = \beta$.

The dual of a N-P function β is the function b on [0,1] given by:

$$b(\lambda) \equiv_\lambda \min_\alpha[(1-\lambda)\alpha + \lambda(1-\beta(\alpha)]$$

More generally if γ is any N-P function such that $\gamma(\alpha) \leqq \beta(\alpha|P_0, P_1)$ for all $\alpha \in [0,1]$ then there is a right continuous monotonically increasing family of test functions $\varphi_\alpha : \alpha \in [0,1]$ in $\mathcal{D} = (P_0, P_1)$ such that

$$E_0\varphi_\alpha \equiv \alpha \quad \textit{while} \quad E_1\varphi_\alpha \equiv \gamma(\alpha).$$

If e.g. γ is given as the upper boundary of the convex hull of points $(0,b), (p_1,q_1), (p_2,q_2)$ and (1,1) where $0 \leqq p_1 \leqq p_2 \leqq 1$ and $\gamma(0) = b, \gamma(p_i) = q_i; i = 1,2$ then we may construct the family $\varphi_\alpha : \alpha \in [0,1]$ in the following steps:

(i) Let $\delta_\alpha : \alpha \in [0,1]$ be given as above.

(ii) Put $\varphi_0 = [b/\beta(0|P_0,P_1)]\delta_0$.

(iii) Let α_1 be the smallest number $\alpha_1 \geqq 0$ such that the graph of $\beta(\cdot|P_0,P_1)$ intersects the line through (0,b) and (p_1,q_1) in the point $(\alpha_1, \beta(\alpha_1|P_0,P_1))$. Put so $\varphi_\alpha = (1-\theta)\varphi_0 + \theta\delta_{\alpha_1}$ for $\alpha = (1-\theta)0 + \theta\alpha_1$ in $[0,p_1]$.

(iv) Let α_2 be the smallest number $\alpha_2 \geqq \alpha_1$ such that the line through (p_1,q_1) and (p_2,q_2) intersects the graph of $\beta(\cdot|P_0,P_1)$ in $(\alpha_2, \beta(\alpha_2|P_0,P_1))$. Put so $\varphi_\alpha = (1-\theta)\varphi_{p_1} + \theta\delta_{\alpha_2}$ for $\alpha = (1-\theta)p_1 + \theta\alpha_2$ in $[p_1,p_2]$.

(v) Put $\varphi_\alpha = (1-\theta)\varphi_{p_2} + \theta\cdot 1$ for $\alpha = (1-\theta)p_2 + \theta \cdot 1$ in $[p_2, 1]$.

It may be checked that $\delta_0 \geqq \varphi_0, \delta_{\alpha_1} \geqq \varphi_{p_1}, \delta_{\alpha_2} \geqq \varphi_{p_2}$ and that $\varphi_0 \leqq \varphi_{p_1} \leqq \varphi_{p_2} \leqq 1$ so that $\varphi_\alpha : \alpha \in [0,1]$ satisfy our requirements.

Proceeding by induction we obtain for any polygonal $\gamma \leqq \beta(\cdot|P_0,P_1)$ a representation $\varphi_\alpha : \alpha \in [0,1]$. By compactness this extends to any N-P function $\gamma \leqq \beta(\cdot|P_0,P_1)$.

This procedure is closely related to the procedure known from the theory of majorization, see e.g. Marshall and Olkin (1979), whereby we may pass from a vector p to a vector q which is majorized by p by a finite number of "decreasing" steps each modifying only two coordinates.

Suppose now that $\gamma(\cdot|Q_0,Q_1) \leqq \beta(\cdot|P_0,P_1)$ for a dichotomy (Q_0,Q_1). Then $\gamma(\alpha|Q_0,Q_1) \equiv E_1\varphi_\alpha$ where $\alpha \equiv E_0\varphi_\alpha$ for an increasing right continuous family $\varphi_\alpha : 0 \leqq \alpha \leqq 1$ of testfunctions in $\mathcal{D} = (P_0,P_1)$. Let $M(\cdot|x)$, for each x in the sample space of $\mathcal{D}$ be the measure on [0,1] having distribution function $\alpha \to \varphi_\alpha(x)$. Letting R(0,1) denote the uniform distribution on (0,1) we find for any Borel set $B \subseteqq [0,1]$ that:

$$R(0,1)(B) = \int M(B|x)P_0(dx)$$

while:

$$\gamma(B|P_0,P_1) = \int_B \gamma(dx|P_0,P_1) = \int M(B|x)P_1(dx)$$

Remarks

The equivalent conditions (i) - (v) express all that $\mathcal{D}$ is at least as informative as $\tilde{\mathcal{D}}$. A dilation on $[0,\infty[$ is a Markov kernel D from $[0,\infty[$ to $[0,\infty[$ such that $\int xD(dx|y) = y; y \geqq 0$.

The integral $\int(dP_1/dP_0)^t dP_0$ for a dichotomy $\mathcal{D} = (P_0, P_1)$ is, as a function of $t \in [0,1]$, the *Hellinger transform* of $\mathcal{D}$. It defines $\mathcal{D}$ up to equivalence. The ordering described by (vi) does not however, see Torgersen (1970), imply that $\mathcal{D}$ is at least as informative as $\tilde{\mathcal{D}}$. In terms of the N-P function β of $\mathcal{D}$ the Hellinger transform may be expressed as:

$$t \to \int_0^1 [\beta'(\alpha)]^t d\alpha = \int_0^1 [K^{-1}(\alpha)]^t d\alpha$$

where

$$K = \mathcal{L}(dP_1/dP_0|P_0).$$

Consider so comparison of dichotomies $\mathcal{D} = (F_0, F_1)$ and $\tilde{\mathcal{D}} = (G_0, G_1)$ where all four distributions are on the real line. Denote the N-P functions of $\mathcal{D}$ and $\tilde{\mathcal{D}}$ by, respectively, β and $\tilde{\beta}$. Let us agree that p-fractiles with $p > 0$ are chosen minimal while 0-fractiles are chosen maximal.

Assume first that $\mathcal{D} \geqq \tilde{\mathcal{D}}$ and that $\mathcal{D}$ has monotonically increasing likelihood ratio. Then, for any number c, $F_0^{-1}(G_0(c)) \leqq F_1^{-1}(G_1(c))$. Indeed if this was not so for a number c then there is a t so that

$$F_1^{-1}(G_1(c)) < t < F_0^{-1}(G_0(c)).$$

Putting $\alpha = 1 - G_0(c))$ we find that $\tilde{\beta}(\alpha) \geqq 1 - G_1(c) \geqq 1 - F_1(t) = \beta(\tilde{\alpha})$ where $\tilde{\alpha} = 1 - F_0(t) \geqq 1 - G_0(c) = \alpha$. Hence $\tilde{\beta}(\alpha) \geqq \beta(\tilde{\alpha}) \geqq \beta(\alpha) \geqq \tilde{\beta}(\alpha)$ so that " $=$" prevails and thus $G_1(c) = F_1(t)$. The convention concerning 0-fractiles implies that $G_1(c) = F_1(t) > 0$ and thus that $\beta(\tilde{\alpha}) = \beta(\alpha) < 1$. Hence, since $\tilde{\alpha} \geqq \alpha$, $\tilde{\alpha} = \alpha$ so that $G_0(c) = F_0(t)$. In particular t is a $G_0(c)$ fractile of F_0. If $G_0(c) > 0$ then this is incompatible with the assumption that $t < F_0^{-1}(G_0(c))$. Thus $G_0(c) = 0$ so that $\alpha = 1$, but this is incompatible with the inequality $\beta(\alpha) < 1$. It follows that there is no number t with the asserted property.

Can this be reversed? It is claimed in Lehmann (1988) that this is permissible if both dichotomies have monotonically increasing likelihood ratio. There is however a missing link in the proof and in fact counterexamples exist when we permit the distributions F_0 and F_1 to have point masses. Assume e.g. that $F_i; i = 0,1$ have masses $1 - p_i$ and p_i in, respectively, 0 and 1 where $p_0 < p_1$. Let γ be a N-P function such that $\gamma(0) = 0$. Then we may let G_0 be the uniform distribution on [0,1] and let G_1 be the distribution on [0,1] having distribution function $G_1(\alpha) \equiv 1 - \gamma(1 - \alpha)$. Then both dichotomies have

As mentioned earlier the N-P functions and their close relatives appear in abundance in statistics and in econometry. Thus spread in econometry is frequently described, see e.g. Arnold (1987), in terms of Lorenz functions. Then relationship between Lorenz functions and N-P functions may be described as follows.

Let F be any distribution on $[0,\infty[$ having finite positive expectation $\mu_F = \int xF(dx) = \int_0^1 F^{-1}(p)dp$. The Lorenz function of F is the function L_F on [0,1] defined by:

$$L_F(p) \equiv_p \int_0^p F^{-1}(t)dt/\mu_F.$$

Put $F_0 = F$ and let F_1 have density $x \to x/\mu_F$ w.r.t. F_0. Let K be the distribution of dF_1/dF_0 under F_0. Then $K^{-1} = F^{-1}/\mu_F$ and $L_F(p) \equiv_p 1-\beta(1-p|\mathcal{D}_F)$ where $\mathcal{D}_F$ in the dichotomy (F_0, F_1).

It is easily inferred that a function L is a Lorenz function if and only if it is of the form $L(\alpha) \equiv 1-\beta(1-\alpha)$ for a N-P function β such that $\beta(0) = 0$. If L is in this form with β being the N-P function of the dichotomy (P_0, P_1) then L is the inverse function of the N-P function of the reversed dichotomy (P_1, P_0).

It follows that a function is a Lorenz function if and only if it is a continuous convex function from [0,1] onto [0,1] having the origin as a fixed point.

Considering two probability distributions F and G on $[0,\infty[$ such that both have finite positive expectations we may, following Arnold (1987), say that G Lorenz majorizes F if $L_F \geqq L_G$. By the above comments this amounts to the condition that $\mathcal{D}_G \geqq \mathcal{D}_F$ where $\mathcal{D}_G$ is defined in terms of G as $\mathcal{D}_F$ above was defined in terms of F.

Another notion related to the N-P functions is the total time on test (TTT) transform in reliability theory. These are, see Klefsjø (1984), the functions of the form $\alpha \to 1-\beta(1-\alpha)+(1-\alpha)\beta'(1-\alpha)$ for a N-P function β. This is particularily relevant here since, as we shall see, reliability theory provides answers to many problems concerning how selection influences information.

where

$$N(\theta_0, h) = \{\theta : \sum_{i=1}^{k} |\theta^i - \theta_0^i| \leqq h\}.$$

If the local defiency $\delta^{\bullet}_{\theta_0}(\mathcal{E}, \mathcal{F}) = 0$ then we shall say that $\mathcal{E}$ *is locally at least as informative as* $\mathcal{F}$ *at* θ_0. *The local deficiency distance* $\Delta^{\bullet}_{\theta_0}(\mathcal{E}, \mathcal{F})$ *between* $\mathcal{E}$ *and* $\mathcal{F}$ *at* $\theta = \theta_0$ may then be introduced as the larger one of the numbers $\delta^{\bullet}_{\theta_0}(\mathcal{E}, \mathcal{F})$ and $\delta^{\bullet}_{\theta_0}(\mathcal{F}, \mathcal{E})$. This local distance behaves then as a pseudometric, just as the "global" deficiency distance $(\mathcal{E}, \mathcal{F}) \to \max\{\delta(\mathcal{E}, \mathcal{F}), \delta(\mathcal{F}, \mathcal{E})\}$ of LeCam (1964 and 1986).

If $\Delta^{\bullet}_{\theta_0}(\mathcal{E}, \mathcal{F}) = 0$ then we shall say that $\mathcal{E}$ and $\mathcal{F}$ are locally equivalent at $\theta = \theta_0$. Of particular relevance here is the fact that in the one-dimensional case an experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ which is differentiable at $\theta = \theta_0$ is, provided $P^{\bullet}_{\theta_0} \neq 0$, locally equivalent with a unique strongly unimodal translation experiment of distributions on the real line.

In general the local properties of $\mathcal{E} = (P_\theta : \theta \in \Theta)$ are stored in the distribution $F(\cdot|\theta_0, \mathcal{E})$ of $(dP^{\bullet}_{\theta_0}, /dP_{\theta_0}; i = 1, \cdots, k)$ under P_{θ_0}. Thus the Fisher information matrix $I(\theta_0, \mathcal{E})$ of $\mathcal{E}$ at θ_0 has as its (i,j)th entry the number $\int x^i x^j F(dx|\theta_0, \mathcal{E})$. Furthermore $\mathcal{E}$ is locally at least as informative as $\mathcal{F}$ at $\theta = \theta_0$ if and only if $F(\cdot|\theta_0, \mathcal{E})$ is a dilation of $F(\cdot|\theta_0, \mathcal{F})$.

In general the ϵ-version of the transition criterion has a dual version expressible in terms of approximate dilations. The distribution $F(\cdot|\theta_0, \mathcal{E})$ has expectation (vector) zero and, conversely, any distribution on R^k having expectation zero qualifies as a distribution $F(\cdot|\theta_0, \mathcal{E})$ for a suitably chosen differentiable experiment $\mathcal{E}$.

In the one dimensional case local comparison may be completely described in terms of another family of concave functions related to the Neyman-Pearson lemma. These are the continuous concave functions κ on the unit interval [0,1] such that $\kappa(0) = \kappa(1) = 0$. A function having these properties is here called a *(N-P) slope function.* Thus a N-P slope function is a concave function on [0,1] such that

$$\kappa(0) = \kappa(0+) = \kappa(1-) = \kappa(1) = 0.$$

Assuming that $\Theta \subseteqq R$ consider an experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ which is differentiable at $\theta_0 \in \Theta$. Consider also a size α-test δ for testing "$\theta = \theta_0$" against "$\theta > \theta_0$". Then $E_{\theta_0}\delta = \alpha$ and the slope of the powerfunction at $\theta = \theta_0$ is

$$\frac{d}{d\theta} E_\theta \delta]_{\theta_0} = \int \delta dP^{\bullet}_{\theta_0}.$$

This slope is maximized by any size α test δ admitting a constant c such that $\delta = 0$ or $\delta = 1$ as $dP^{\bullet}_{\theta_0}/dP_{\theta_0} < c$ or $dP^{\bullet}_{\theta_0}/dP_{\theta_0} > c$. Insisting that the test should be functionally dependent on $dP^{\bullet}_{\theta_0}/dP_{\theta_0}$ we obtain, for each $\alpha \in [0,1]$, a unique maximizing test δ.

The basic observation here is that this maximum slope as a function of the size $\alpha \in [0,1]$ is a N-P slope function which we here shall denote by $\kappa(\cdot|\theta_0, \mathcal{E})$. Any N-P slope function assigns masses $B/(A+B)$ and $A/(A+B)$ to, respectively -A and B.

Putting $M = AB/(A+B)$ and $p = A/(A+B)$ we obtain all non null triangular N-P slope functions.

A translation experiment $\mathcal{E} = (G_\theta : \theta \in R)$ on the real line is differentiable if and only if G has an absolutely continuous density g such that $\int_{-\infty}^{+\infty} |g'(x)|dx < \infty$. In that case the slope function $\kappa = \kappa(\cdot|\theta_0, \mathcal{E})$ does not depend on θ_0. If G is strongly unimodal then κ may be expressed as $\kappa(\alpha) \equiv g(G^{-1}(1-\alpha))$.

Conversely if κ is a non null N-P slope function then κ is of this form for a strongly unimodal distribution function G such that the corresponding translation experiment is differentiable. This family of distribution functions constitutes the totality of solutions of the differential equation $G' = \kappa(1-G)$.

In the one dimensional case all local properties may be expressed in terms of the slope function. Thus if $\mathcal{E}$ has slope function κ at θ_0 then the Fisher information at θ_0 is the quantity

$$I(\theta_0, \mathcal{E}) = \int x^2 F(dx|\theta_0, \mathcal{E}) = \int_0^1 \kappa'(\alpha)^2 d\alpha = \int_0^1 [\frac{\kappa(\alpha)}{\alpha} - \kappa'(\alpha)]^2 d\alpha.$$

The slope function κ may be expressed directly in terms of $F = F(\cdot|\theta_0, \mathcal{E})$ by:

$$\kappa(\alpha) \equiv_\alpha \int_0^\alpha F^{-1}(1-p)dp$$

so that the lower $1-\alpha$ fractile $F^{-1}(1-\alpha)$ of F is the right derivative of κ for any $\alpha \in]0,1[$.

Consider now an everywhere differentiable experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ having an open interval of the real line as its parameter set. Then any test function δ in $\mathcal{E}$ satisfies the differential inequality

$$\frac{d}{d\theta} E_\theta \delta \leqq \kappa(E_\theta \delta | \theta, \mathcal{E}); \theta \in \Theta$$

with "=" everywhere if δ is a most powerful level $E_{\theta_1}\delta$ test for testing " θ_1" against "θ_2" whenever $\theta_1 < \theta_2$.

The last observation is crucial for linking local and the global information in differentiable M-L experiments.

Proof:
Let us, as the "if" is trivial, assume that $\mathcal{E} = (P_\theta : \theta \in \Theta)$ is pairwise more informative than $\mathcal{F} = (Q_\theta : \theta \in \Theta)$

Let σ be a powerfunction in $\mathcal{F}$ and let $\theta_0 \in \Theta$. Put $\alpha = \sigma(\theta_0)$ and assume first that $0 < \alpha < 1$. By the M-L property of $\mathcal{E}$ there is a powerfunction π in $\prod_{\mathcal{E}}$ so that $\pi(\theta_0) = \alpha$. If $\theta \geqq \theta_0$ then

$$\pi(\theta) = \beta(\alpha|P_{\theta_0}, P_\theta) \geqq \beta(\alpha|Q_{\theta_0}, Q_\theta) \geqq \sigma(\theta).$$

If $\theta \leqq \theta_0$ then

$$\beta(\pi(\theta)|P_{\theta_1}, P_{\theta_0}) = \alpha \leqq \beta(\sigma(\theta)|Q_{\theta_1}, Q_{\theta_0}) \leqq \beta(\sigma(\theta)|P_{\theta_1}, P_{\theta_0})$$

so that $\pi(\theta) \leqq \sigma(\theta)$.

Consider so the case where $\sigma(\theta_0) = 0$ or $\sigma(\theta_0) = 1$. Replace σ with $\sigma_n = (1 - \frac{1}{n})\sigma + \frac{1}{n}$ or by $\sigma_n = (1 - \frac{1}{n})\sigma$ as, respectively, $\sigma(\theta_0) = 0$ or $\sigma(\theta_0) = 1$.

Applying the above result to σ_n with $n \geqq 2$ and then using that $\prod_{\mathcal{E}}$ is dosed for pointwise convergence we obtain a powerfunction π in $\pi_{\mathcal{E}}$ so that $\pi(\theta) \leqq \sigma(\theta)$ or $\pi(\theta) \geqq \sigma(\theta)$ as $\theta \leqq \theta_0$ or $\theta \geqq \theta_0$. The last statement follows from the fact that if $\mathcal{F}$ has the M-L property and if σ is a powerfunction in $\mathcal{F}$ and if $\theta_0 \in \Theta$ then there is powerfunction $\hat{\sigma}$ in $\prod_{\mathcal{F}}$ so that $\hat{\sigma}(\theta) \leqq \sigma(\theta)$ or $\hat{\sigma}(\theta) \geqq \sigma(\theta)$ as $\theta \leqq \theta_0$ or $\theta \geqq \theta_0$.

□

Corollary 4.2

M-L experiments $\mathcal{E}$ and $\mathcal{F}$ are pairwise equivalent if and only if $\prod_{\mathcal{E}} = \prod_{\mathcal{F}}$.

Remark
We shall soon see that pairwise equivalence and full equivalence is the same thing for M-L experiments.

With the modifications described in section 2 the criterion og Lehmann (1988) applies:

Theorem 4.3. (Pairwise comparability and distribution functions)

Let $\mathcal{E} = (F_\theta : \theta \in \Theta)$ and $\mathcal{F} = (G_\theta : \theta \in \Theta)$ be experiments on the real line. Then:

(a) If $\mathcal{E}$ has monotonically increasing likelihood ratio in $T(x) \equiv x$ and if $\mathcal{E} \geqq \mathcal{F}$ pairwise then $F_\theta^{-1}(G_\theta(x))$ is monotonically increasing in θ for any x.

(b) If $\mathcal{F}$ has monotonically increasing likelihood ration in $T(x) \equiv x$ and if the distributions in $\mathcal{E}$ are non atomic then $\mathcal{E} \geqq \mathcal{F}$ pairwise provided $F_\theta^{-1}(G_\theta(x))$ is monotonically increasing in θ for any x.

monotone decison problems. This may be seen directly from Lehmann's criterion since $H^{-1}_{\theta,\sigma_1}(H_{\theta,\sigma_2}(x)) = (1-\frac{\sigma_1}{\sigma_2})\theta + \frac{\sigma_1}{\sigma_2}x$ when $0 < H_{\theta,\sigma_2}(x) < 1$.

We shall see in section 6 that if an experiment $\mathcal{E}$ is at least as informative as another experiment $\mathcal{F}$ *for a given loss function* then $\mathcal{E}^n$ is at least as informative as $\mathcal{F}^n$ for the same loss function. Thus theorem 4.1, and also theorem 4.3 when it applies, provides conditions ensuring that n independent observations of a variable X are more informative than n independent observations of a variable Y pairwise as well as for monotone decision problems. In particular this yields the n-sample results in Lehmann (1988). It should however be emphasized that the criteria ensuring that $\mathcal{E}^n \geqq \mathcal{F}^n$ for monotone decision problems are not the same as those ensuring that $\mathcal{E} \geqq \mathcal{F}$ for monotone decision problems. This is so even when both $\mathcal{E}$ and $\mathcal{F}$ are assumed to possess the M-L property. Indeed any pair $(\mathcal{E},\mathcal{F})$ of non comparable dichotomies such that $\mathcal{E}^n \geqq \mathcal{F}^n$ provides a counterexample.

Let us take a closer look on the sets $\prod_\mathcal{E}$ of powerfunctions. For any experiment $\mathcal{E}$ (with $\Theta \subseteqq R$) the set $\prod_\mathcal{E}$ is a family of monotonically increasing functions from Θ to [0,1] having the additional properties:

(a') If $\theta_0 \in \Theta$ and $0 < \alpha < 1$ then there is at most one function π in $\prod_\mathcal{E}$ such that $\pi(\theta_0) = \alpha$.
(b) To each pair (θ_1,θ_2) of points in Θ such that $\theta_1 < \theta_2$ corresponds a N-P function $\beta(\cdot|\theta_1,\theta_2)$ such that $\pi(\theta_2) = \beta(\pi(\theta_1)|\theta_1,\theta_2)$ whenever $\pi(\theta_1) > 0$.
(c) $\prod_\mathcal{E}$ is closed for pointwise convergence on Θ.

Conditions (a') and (b) imply together that $\prod_\mathcal{E}$ is totally ordered for the pointwise ordering. If $\mathcal{E}$ possesses monotone likelihood ratio then (a') may be strengthened to:

(a) If $\theta_0 \in \Theta$ and $0 < \alpha < 1$ then there is a unique powerfunction π in $\prod_\mathcal{E}$ so that $\pi(\theta_0) = \alpha$.

Condition (c) is not as imposing as it may appear. Actually if $\prod$ is a set of monotonically increasing functions on Θ satisfying (a) and (b), then (c) amounts to the condition that $\prod$ should contain those indicator functions which are pointwise limits of functions from $\prod$.

We shall later see that conditions (a) and (b) together characterize the M-L property.

For now observe the following facts concerning the tests having a particular function function $\pi \in \prod_\mathcal{E}$ as its powerfunction:

If $\pi \in \prod_\mathcal{E}$ and δ is a test such that $E_{\theta_0}\delta \leqq \pi(\theta_0) > 0$ then $E_\theta\delta \leqq \pi(\theta)$ when $\theta \geqq \theta_0$.

If $\pi \in \prod_\mathcal{E}$ and δ is a test such that $E_{\theta_1}\delta \geqq \pi(\theta_1) < 1$ then $E_\theta\delta \geqq \pi(\theta)$ when $\theta \leqq \theta_1$.

Joining these "principles" we conclude that if $E_\theta\delta = \pi(\theta)$ for $\theta = \theta_0$ and for $\theta = \theta_1$ then this is so also for $\theta_0 \leqq \theta \leqq \theta_1$.

It follows that if $\mathcal{E}$ is homogeneous and $\theta_0 \leqq \theta \leqq \theta_1$ for any $\theta \in \Theta$ and if $\pi \in \prod_\mathcal{E}$ then $E_\theta\delta_\pi \equiv_\theta \pi(\theta)$ where δ_π is the unique most powerful level $\pi(\theta_0)$ test for testing

If $P_{\theta_0}(C) > 0$ then the requirement may be written:

$$E_{\theta_0}(\delta_\pi|C) = E_{\theta_0}(\phi|C)(= \frac{E_{\theta_0}\phi - P_{\theta_0}(B)}{P_{\theta_0}(C)}).$$

Varying k and γ the quantity $E_{\theta_0}(\delta_\pi|C)$ varies freely in [0,1]. Thus an assignment such that $E_{\theta_0}\delta_\pi = E_{\theta_0}\phi = \pi(\theta_0)$ is feasible. By the Neyman-Pearson lemma $E_{\theta_1}(\delta_\pi|C) \geqq E_{\theta_1}(\phi|C)$ when $P_{\theta_1}(C) > 0$. In any case $\int_C \delta_\pi dP_{\theta_1} \geqq \int_C \phi dP_{\theta_1}$ so that $E_{\theta_1}\delta_\pi \geqq E_{\theta_1}\phi$. Thus, by optimality, $E_{\theta_1}\delta_\pi = E_{\theta_1}\phi = \pi(\theta_1)$. Hence $E_\theta\delta_\pi = E_\theta\phi = \pi(\theta); \theta_0 \leqq \theta \leqq \theta_1$ so that $E_\theta\delta_\pi \equiv \pi(\theta)$.

Furthermore by the uniqueness part of the Neyman-Pearson lemma δ_π is, up to nullsets, the unique, test which has powerfunction π and which is measurable w.r.t. the minimal sufficient σ-algebra. The above discussion provides a substantial part of the spade work for:

Theorem 4.4 (Monotone assignment of tests 1).

If Θ is finite then there corresponds to any given powerfunction π in $\prod_\mathcal{E}$ one and only one test-function δ_π having that powerfunction and which is measurable w.r.t. the minimal sufficient σ-algebra. This test-function δ_π assumes, disregarding nullsets, besides 0 and 1 at most one additional value in]0,1[. Furthermore the assignment $\pi \to \delta_\pi$ is monotonically increasing.

Proof:
We may without loss of generality assume that $\Theta = \{1, \cdots, m\}$. Put $\mu = \sum_\theta P_\theta$ and $f_\theta = dP_\theta/d\mu$. We supress the qualification a.e. in the arguments below.

Let $\prod$ be the set of functions in $\prod_\mathcal{E}$ which are not indicator functions. Obviously we may restrict attention to powerfunctions in $\prod$.

Decompose $\prod$ as $\prod = \bigcup_{i=1}^m \prod^i$ where $\prod^i = \{\pi : \pi \in \prod, \pi(i) > 0, \pi(\theta) = 0 \text{ when } \theta < i\}$
Then, since $\prod$ is totally ordered, any power function in $\prod^i$ majorizes any powerfunction in $\prod^k$ when $i < k$ i.e., with obvious notations,:

$$\prod^1 \geqq \prod^2 \geqq \quad \geqq \prod^m.$$

Decompose so each set $\prod^i$ as

$$\prod^i = \bigcup_{i=1}^m \prod^{i,j}$$

There is for any experiment $\mathcal{E}$ a monotone assignment $\pi \rightarrow \delta_\pi$ from $\prod_\mathcal{E}$ to the set of test functions such that

$$E_\theta\delta_\pi \equiv_\theta \pi(\theta) \text{ for all} \pi \in \prod_\mathcal{E}$$

Remark.
If the conclusion of the weak compactness lemma is not valid for $\mathcal{E}$ then the statement of the theorem requires that we admit the "generalized" test functions described at the beginning of this section.

An important and immediate consequence is:

Corollary 4.6 (Powerdiagrams and types of M-L experiments)

If $\prod_\mathcal{E} = \prod_\mathcal{F}$ for experiments $\mathcal{E}$ and $\mathcal{F}$ and if $\mathcal{F}$ has the monotone likelihood property then the experiments $\mathcal{E}$ and $\mathcal{F}$ are equivalent.

Proof:
Express $\mathcal{E}$ and $\mathcal{F}$ as $\mathcal{E} = (P_\theta : \theta \in \Theta)$ and $\mathcal{F} = (Q_\theta : \theta \in \Theta)$. It follows, since $\mathcal{F}$ has monotone likelihood, that the set $\prod = \prod_\mathcal{E} = \prod_\mathcal{F}$ satisfies conditions (a) and (b). Thus $\mathcal{E}$ and $\mathcal{F}$ are pairwise equivalent.

Associate with each constant $c \in R$ the test in $\mathcal{F}$ with rejection region $[T > c]$. Then π^c given by $\pi^c(\theta) \equiv_\theta Q_\theta(T > c)$ is in $\prod$. It follows that there is a test function δ^c in $\mathcal{E}$ such that $\pi^c(\theta) \equiv_\theta E_\theta\delta_c$ where $\delta_{c_1} \leqq \delta_{c_2}$ a.e. $\mathcal{E}$ when $c_1 \geqq c_2$. In order to escape technicalities we may, and shall, assume that Θ is finite. Then we may regularize the map $c \rightarrow \delta_c$ so that it is pointwise monotonically decreasing and continuous from the right and such that $\delta_c \rightarrow 0$ or $\rightarrow 1$ as $c \rightarrow \infty$ or $c \rightarrow -\infty$. Then there is a unique Markov kernel M from $\mathcal{E}$ to $\mathcal{F}$ such that $M(]-\infty, c]|\cdot) = 1 - \delta_c$ for all numbers c. Clearly $E_\theta M(]-\infty, c]|\cdot) = 1 - \pi^c(\theta) = Q_\theta(T \leqq c)$ so that $P_\theta M \equiv_\theta \mathcal{L}(T|\theta)$. It follows, since T is sufficient in $\mathcal{F}$, that $\mathcal{E}$ is at least as informative as $\mathcal{F}$. Hence, since they are pairwise equivalent, they are equivalent.

□

We shall now see that the M-L property is a property of type i.e. that any experiment which is statistically equivalent with a M-L experiment is itself a M-L experiment. By the last corollary this implies that if $\prod_\mathcal{E} = \prod_\mathcal{F}$ for a M-L experiment $\mathcal{F}$ then also $\mathcal{E}$ has the M-L property.

Let $\mathcal{E} = (P_\theta : \theta \in \Theta)$ be an experiment having monotonically increasing likelihood ratio in a statistic T. If f_θ is the density of P_θ w.r.t. some majorizing measure μ and if $\theta_1 \leqq \theta_2$

Assume that $u\underset{=}{]}v\underset{=}{]}u$ where both functions are non null. If $v_{\theta_0} > 0$ while $u_{\theta_0} = 0$ then $v_\theta/v_{\theta_0} \geqq u_\theta/u_{\theta_0}$ when $\theta > \theta_0$ so that $u_\theta = 0$ when $\theta \geqq \theta_0$. Hence, since $u \neq 0, u_{\theta_1} > 0$ for some $\theta_1 < \theta_0$. Then $0 = u_{\theta_0}/u_{\theta_1} \geqq v_{\theta_0}/v_{\theta_1}$ contradicting the assumption that $v_{\theta_0} > 0$. Thus $v_\theta > 0$ if and only if $u_\theta > 0$.

Choose a parameter value θ_0 such that $v_{\theta_0} > 0$ and put $t = u_{\theta_0}/v_{\theta_0}$. If $\theta > \theta_0$ and $v_\theta > 0$ then $v_\theta/v_{\theta_0} = u_\theta/u_{\theta_0}$ so that $u_\theta/v_\theta = t$. Likewise if $\theta < \theta_0$ and $v_\theta > 0$ then $v_{\theta_0}/v_\theta = u_{\theta_0}/u_\theta$ so that again $u_\theta/v_\theta = t$.

□

In general if V is a candidate for the set possible likelihood functions then there should to each θ in Θ be at least one function v in V such that $v_\theta > 0$. Assuming this we may draw the conclusion that if V is also totally ordered for $\underset{=}{]}$ then zeros and positive values of a function in V appears along the θ axis as:

$$0,\cdots,0,+,\cdots,+,0,\cdots,0$$

where one or both sequences of zeros may be empty.

In fact if $v \in V$ and if $v_{\theta_2} = 0$ while $v_{\theta_1} > 0$ and $v_{\theta_3} > 0$ where $\theta_1 < \theta_2 < \theta_3$ then we find for any $w \in V$ with $w_{\theta_2} > 0$ that $w_{\theta_2}/w_{\theta_1} > 0 = v_{\theta_2}/v_{\theta_1}$ while $w_{\theta_3}/w_{\theta_2} < \infty = v_{\theta_3}/v_{\theta_2}$. This, however, contradicts either of the inequalities $w\underset{=}{[}v$ and $v\underset{=}{[}w$.

Let us agree to say that a non negative function v on Θ is *sign regular* if $\{\theta : v_\theta > 0\}$ is a subinterval of Θ. In other words v is sign regular if $v_\theta > 0$ whenever $\theta_1 < \theta < \theta_2$ where $v_{\theta_1} > 0$ and $v_{\theta_2} > 0$.

It may now be checked that if $v\underset{=}{]}w$ where v and w are both sign regular then passing from w to v the end points of the interval of positivity remain either unchanged or are pushed to the right.

Sets of likelihood functions which are totally ordered for the relation $\underset{=}{]}$ need of course not be measurable. If Θ is finite however then maximal sets having this property are necessarily measurable.

In fact, whether Θ is finite or not, if V is a set of likelihood functions which is totally ordered for the relation $\underset{=}{]}$ then, by proposition 4.7, also the pointwise closure of V is totally ordered for this relation. If Θ is finite then closed subsets of R^Θ are measurable.

Say that a real valued function T on V is strictly increasing if $T(v) \leqq T(w)$ whenever $v\underset{=}{[}w$ while $T(v) = T(w)$ if and only if v and w are positivily proportional.

Totally ordered sets V may be represented by subsets of the real line with the usual ordering as follows:

Remark

The *standard measure* of $\mathcal{E} = (P_\theta : \theta \in \Theta)$ is the measure μf^{-1} where $\mu = \sum_\theta P_\theta, f_\theta = dP_\theta/d\mu; \theta \in \Theta$ and $f = (f_\theta : \theta \in \Theta)$. By Blackwell (1951) and LeCam (1964) this measure characterizes $\mathcal{E}$ up to equivalence. The experiment $\mathcal{E}f^{-1} = (P_\theta f^{-1} : \theta \in \Theta)$ is called *the standard experiment of* $\mathcal{E}$.

Proof:

We shall use the notations of the remark. If $\mathcal{E}$ has the M-L property then we noted above that, disregarding a $\mathcal{E}$ null set, the set of possible functions $\theta \to f_\theta$ is totally ordered for the ordering $\underset{=}{]}$. Hence μf^{-1} has a totally ordered support. Conversely if μf^{-1} has a totally ordered support then $f_{\theta_2}/f_{\theta_1} = \phi_{\theta_2,\theta_1}(T(f))$ for ϕ_{θ_2,θ_1} and T as constructed above.

□

Corollary 4.10 (Types of M-L experiments contain M-L experiments only)

An experiment which is equivalent with a M-L experiment is itself a M-L experiment.

Remark.

It suffices actually to require pairwise equivalence. Indeed, as we soon shall see, the pairwise equivalence of $\mathcal{E}$ and a M-L experiment implies full equivalence.

Proof:

This follows directly from the theorem and the definition of the M-L property.

□

In order to show that the monotone likelihood property is a triplewise property we shall need:

Proposition 4.11

Assume $\Theta = \{1, \cdots, m\}$ and that v and w are non negative functions on Θ such that

$$(v_i, v_{i+1})\underset{=}{]}(w_i, w_{i+1}); i = 1, \cdots, m-1.$$

Assume also that v and w are both sign regular. Then either $v\underset{=}{]}w$ or v and w may be expressed as:

$$v = (v_1, \cdots, v_{t-1}, 0, \cdots, 0), w = (0, \cdots, 0, w_{t+1}, \cdots, w_m)$$

supports the standard measure of $\mathcal{E}|\Theta_i$. Then the standard measure of $\mathcal{E}$ is supported by the set V of likelihood functions v on Θ such that any restriction $v|\Theta_i$ is in $W_i; i = 1, \cdots, 4$. The proof will now be completed by showing that V is totally ordered. Consider functions v and w i V. We must show that either $v \underset{=}{[} w$ or $v \underset{=}{]} w$.

Assume that neither $v \underset{=}{[} w$ nor $v \underset{=}{]} w$. Note that our assumptions imply that v and w are both sign regular. It follows that there are indices $k, k+1$ and $i, i+1$ such that $w_k > 0$ and $v_{k+1}/v_k > w_{k+1}/w_k$ while $v_i > 0$ and $v_{i+1}/v_i < w_{i+1}/w_i$. We may without loss of generality assume that $k < i$ and then, since $\{k, k+1, i, i+1\}$ can't be contained in a three point set, $k = 1$ and $i = 3$. Thus $w_1 > 0$ and $v_2/v_1 > w_2/w_1$ while $v_3 > 0$ and $v_4/v_3 < w_4/w_3$. In particular $v_2 > 0$ and $w_4 > 0$. Our assumptions imply that "$v \underset{=}{]} w$" on the sets $\{1,2,3\}$ and $\{1,2,4\}$ while "$v \underset{=}{[} w$" on the sets $\{2,3,4\}$ and $\{1,3,4\}$. Hence $v_3/v_2 = w_3/w_2, v_3/v_1 = w_3/w_1, v_4/v_2 = w_4/w_2$, and $v_4/v_1 = w_4/w_1$.

Furthermore the quantities $v_2/v_1, w_2/w_1$, and $v_3/v_1 = w_3/w_1$ are positive and finite. This, however, yields the contradiction:

$$w_3/w_1 = v_3/v_1 = v_3/v_2 \cdot v_2/v_1 > w_3/w_2 \cdot w_2/w_1 = w_3/w_1$$

$\square$

Let us return to our considerations on powerfunctions. We have seen that the set $\prod_{\mathcal{E}}$ of powerfunctions characterizes a M-L experiment $\mathcal{E}$ up to equivalence. The question then naturally arises: What sets $\prod$ of function are of the form $\prod = \prod_{\mathcal{E}}$ for some M-L experiment $\mathcal{E}$?

We have seen that any such set $\prod$ is closed and satisfies conditions (a)-(b). This, however, is all we can say since, as we now shall see, any closed set $\prod$ of monotonically increasing functions satisfying these conditions is of the form $\prod_{\mathcal{E}}$ for some M-L experiment $\mathcal{E}$.

Assume then that $\prod$ is a set of monotonically increasing functions on Θ satisfying (a) and (b). Let us also for the moment make the simplifying assumption that the only indicator functions in $\prod$ are the constant functions $\pi = 0$ and $\pi = 1$. (This amounts to require that the corresponding M-L experiment is homogeneous). Then the set $\prod$ is automatically closed.

Choose a point $\theta_0 \in \Theta$ and let $\pi = \pi(\cdot|\theta_0, \alpha)$ be the unique function in $\prod$ such that $\pi(\theta_0) = \alpha$. Note that for each $\theta, \pi(\theta|\theta_0, \alpha)$ is a continuous distribution function F_θ on [0,1]. The proof of the above assertion is completed by checking that $\mathcal{F} = (F_\theta : \theta \in \Theta)$ is a M-L experiment such that $\prod = \prod_{\mathcal{F}}$. In fact F_θ is continuous and it is convex or concave as $\theta \leqq \theta_0$ or $\theta \geqq \theta_0$. In particular F_{θ_0} is the uniform distribution on [0,1]. If $\theta_2 > \theta_1$ then $F_{\theta_2}(\alpha) = \beta(F_{\theta_1}(\alpha)|\theta_1, \theta_2)$ for a N-P function $\beta(\cdot|\theta_1, \theta_2)$ and it may be checked that

$$F'_{\theta_2}(\alpha) = \beta'(F_{\theta_1}(\alpha)|\theta_1, \theta_2) F'_{\theta_1}(\alpha)$$

for almost all α. It follows that $\mathcal{F}$ has monotonically decreasing likelihood ratio in T where $T(\alpha) \equiv_\alpha \alpha$. This implies readily that $\prod = \prod_{\mathcal{F}}$.

Remark.
By Torgersen (1977) m-wise equivalence for any given integer m, however large, does not suffice to ensure complete equivalence.

Proof:
We may without loss of generality assume that Θ is finite. Assume the theorem is proved when Θ is a three point set. Let $\mathcal{E}$ be pairwise equivalent with the M-L experiment $\mathcal{F}$. Then we may conclude for any three point subparameter set Θ' that the restricted experiments $\mathcal{E}|\Theta'$ and $\mathcal{F}|\Theta'$ are equivalent and that $\mathcal{E}|\Theta'$ has the M-L property. Then, by theorem 4.12, the experiment $\mathcal{E}$ has the M-L property. By corollary 4.2 the sets $\prod_{\mathcal{E}}$ and $\prod_{\mathcal{F}}$ coincide and thus, by corollary 4.6, $\mathcal{E}$ and $\mathcal{F}$ are equivalent.

It follows that we without loss of generality may assume that $\Theta = \{1,2,3\}$. Let $\mathcal{E} = (P_1, P_2, P_3)$ be pairwise equivalent with the M-L experiment $\mathcal{F} = (Q_1, Q_2, Q_3)$. It suffices, by corollary 4.6, to show that $\prod_{\mathcal{E}} = \prod_{\mathcal{F}}$ and this, since $\mathcal{F}$ has the M-L property, amounts to the condition that $\prod_{\mathcal{F}} \subseteqq \prod_{\mathcal{E}}$.

Consider a powerfunction π in $\prod_{\mathcal{F}}$. Say that π is representable if it is the powerfunction of a test in $\mathcal{E}$. Our task is then to show that π is representable. This is readily checked when π is one of the at most four possible indicator functions in $\prod_{\mathcal{F}}$. If $0 < \pi(1)$ and $\pi(3) < 1$ then, by pairwise equivalence and by optimality, any most powerful level "$\pi(1)$" test for testing "$\theta = 1$" against "$\theta = 3$" in $\mathcal{E}$ has powerfunction π. Hence π is representable in this case.

Consider next the case where $\pi(1) = 0$ and $\pi(3) = 1$. Put $f_\theta = dP_\theta / d\sum_\theta P_\theta$ and put $\delta = (1-\lambda)I_{f_3>0} + \lambda I_{f_1=0}$. Then, since $P_1 \wedge P_3 = 0, E_\theta\delta = \pi(\theta); \theta = 1,3$. Furthermore $P_2(f_1 = 0) \geqq \int_{f_1=0} \psi f_2 = \int \psi f_2 = \pi(2)$ for any testfunction ψ in $\mathcal{E}$ such that $E_\theta\psi = \pi(\theta); \theta = 1,2$ (Then $\psi = 0$ a.e. when $f_1 > 0$). Similarly

$$P_3(f_3 > 0) = \int_{f_3>0} f_2 = \int_{f_3>0} \phi f_2 \leqq E_2\phi = \pi(2)$$

for any test function ϕ in $\mathcal{E}$ such that $E_\theta\phi = \pi(\theta); \theta = 2,3$. (Then $\phi = 1$ a.e. when $f_3 > 0$). Thus $\lambda \in [0,1]$ may be adjusted so that $E_\theta\delta \equiv \pi(\theta)$.

If $0 = \pi(1)$ and $\pi(3) < 1$ then there is, for $n = 2,3,\cdots$, a powerfunction π_n in $\prod_{\mathcal{F}}$ so that $\pi_n(1) = 1/n$. Then $\pi_n \downarrow \pi^* \in \prod_{\mathcal{F}}$ and clearly $\pi^* \geqq \pi$ while $\pi^*(1) = 0 = \pi(1)$. If $\pi^*(3) = 1$ then π^* is representable by the above argument. If $\pi^*(3) < 1$ then $\pi_n(3) < 1$ for n sufficiently large and then π_n is representable. Thus by the weak compactness lemma, π^* is representable in any case. Let δ^* be a test in $\mathcal{E}$ so that $E_i\delta = \pi(i); i = 1,2,3$. As $\pi^*(2) \geqq \pi(2)$ and $\pi^*(3) \geqq \pi(3)$ we may by proposition 2.5 choose a most powerful level $\pi(2)$ test $\tilde{\delta}$ in $\mathcal{E}$ for testing "$\theta = 2$" against "$\theta = 3$" such that $\tilde{\delta} \leqq \delta^*$. Then $E_i\tilde{\delta} = \pi(i); i = 1,2,3$.

If conversely this condition is satisfied for $P_1 + P_2 + P_3$ almost all x then there corresponds to each positive rational number $r <$ the $P_1 + P_2$ essential supremum of dP_2/dP_1 a non negative number t_r such that $(f_2 - rf_1)(f_3 - t_r f_2) \geqq 0$ a.e. $P_1 + P_2 + P_3$.

Modifying the densities we may ensure that $(f_2 - rf_1)(f_3 - t_r f_2) \geqq 0$ for all such rational numbers r.

Consider points x and y in the sample space of $\mathcal{E}$ such that

$$f(x) = (f_1(x), f_2(x), f_3(x)) \text{ and } f(y) = (f_1(y), f_2(y), f_3(y))$$

are not comparable for the ordering $\underline{\underline{]}}$.

We may without loss of generality assume that $(f_2(x), f_3(x))\underline{\underline{]}}(f_2(y), f_3(y))$. Then, by proposition 4.11, the inequality $(f_1(x), f_2(x))\underline{\underline{]}}(f_1,(y), f_2(y))$ can't hold. Thus $f_1(x) > 0, f_2(y) > 0$ and $f_2(x)/f_1(x) < f_2(y)/f_1(y)$. Let r be a rational number in the interval $]f_2(x)/f_1(x), f_2(y)/f_1(y)[$. Then $f_2(x) < rf_1(x)$ while $f_2(y) > rf_1(y)$. The assumption on signs tells us then that $f_3(x) \leqq t_r f_2(x)$ and that $f_3(y) \geqq t_r f_2(y)$. If $f_2(x) = 0$ then this implies that also $f_3(x) = 0$ and this is excluded by hypothesis since then $f(y)\underline{\underline{[}}f(x)$. Thus $f_2(x) > 0$ and hence $f_3(x)/f_2(x) \leqq t_r \leqq f_3(y)/f_2(y)$. Together with the inequality $(f_2(x), f_3(x))\underline{\underline{]}}(f_2(y), f_3(y))$ this shows that $f_3(x)/f_2(x) = f_3(y)/f_2(y)$ and thus that $f(x)\underline{\underline{[}}f(y)$. As this is also contrary to our hypothesis we are forced to conclude that $f(x)$ and $f(y)$ are comparable for any pair (x, y) of points in the sample space of $\mathcal{E}$.

□

If the experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ has the monotone likelihood property then

$$\beta(\alpha|P_{\theta_1}, P_{\theta_3}) \equiv_\alpha \beta(\beta(\alpha|P_{\theta_1}, P_{\theta_2})|\theta_2, \theta_3)$$

for parameter points θ_1, θ_2 and θ_3 such that $\theta_1 < \theta_2 < \theta_3$. This is the *basic comparison rule* governing the pairwise behaviour of M-L experiments.

Consider conversely a family $\beta_{\theta_1,\theta_2}; \theta_1 < \theta_2$ of N-P functions such that

$$\beta_{\theta_2,\theta_3}(\beta_{\theta_1,\theta_2}) = \beta_{\theta_1,\theta_3}$$

when $\theta_1 < \theta_2 < \theta_3$. Let $\prod$ be the set of functions π from Θ to [0,1] such that:

(i) $\beta(\pi(\theta_1)|\theta_1, \theta_2) = \pi(\theta_2)$ when $\theta_1 < \theta_2$ and $\pi(\theta_1) > 0$

and

(ii) $\pi(\theta) > 0$ whenever there is a $\theta' > \theta$ such that $1 > \pi(\theta') > \beta(0|\theta, \theta')$.

If $\theta_0 \in \Theta$ and $0 < \alpha < 1$ then we may construct a function $\pi \in \prod$ such that $\pi(\theta_0) = \alpha$ by putting:

$$\pi(\theta) = \begin{cases} \beta(\alpha|\theta_0, \theta) & \text{when } \theta > \theta_0 \\ \alpha & \text{when } \theta = \theta_0 \end{cases}$$

Starting with the N-P functions $\overline{\beta}(\cdot|\theta_1,\theta_2) : \theta_1 < \theta_2$ we may associate with each finite subset F of Θ a β function $\overline{\beta}_F$ as follows:

Arrange the distinct numbers in F in increasing order as $F = \{a_0, a_1, \cdots, a_m\}$ and let $\overline{\beta}_F$ denote the composed N-P function $\beta_{a_{m-1},a_m}\beta_{a_{m-2},a_{m-1}}\cdots\beta_{a_0,a_1}$.

By our last inequality $\overline{\beta}_F \leqq \overline{\beta}_G$ if $F \subseteqq G$ and the sets F and G have the same smallest and largest elements. Put finally $\beta^*(\cdot|\theta_1,\theta_2) = \sup \beta_F$ where the sup is taken for all finite sets F having θ_1 and θ_2 as, respectively, its smallest and its largest element. As $\overline{\beta}_F \leqq \overline{\beta}_G$ when $F \subseteqq G$ it does not matter whether sup here is interpreted pointwise or for the informational ordering. If $\theta_1 < \theta_2 < \theta_3$ then the sets F appearing in the definition of $\beta^*(\cdot|\theta_1,\theta_3)$ may be chosen so that they all contain θ_2. It follows readily that the family $(\beta^*(\cdot|\theta_1,\theta_2) : \theta_1 < \theta_2)$ obeys the composition rule. Thus there is a M-L experiment $\mathcal{E}^* = (P^*_\theta : \theta \in \Theta)$ such that

$$\beta^*(\alpha|\theta_1,\theta_2) \equiv_\alpha \beta(\alpha|P^*_{\theta_1}, P^*_{\theta_2}).$$

It is straight forward to check that $\mathcal{E}^* \geqq \mathcal{E}^t$ pairwise for all $t \in T$ and furthermore that $\mathcal{E}^* \geqq \mathcal{F}$ pairwise for any other M-L experiment $\mathcal{F}$ such that $\mathcal{F} \geqq \mathcal{E}^t$ pairwise for all $t \in T$.

Replacing sup with inf we obtain the construction of a greatest lower bound of the family $\mathcal{E}^t : t \in T$. This construction is slightly simpler to interpret since informational infima of N-P functions coincide with the corresponding pointwise infima. We have proved:

Theorem 4.17 (The pairwise ordering of M-L experiments).

The collection of types of M-L experiments is order complete for the pairwise ordering. Smallest upper bounds and greatest lower bounds of families of M-L experiments may be obtained from the N-P functions by the above constructions.

In principle we should now be able to approach the interesting problem of finding for a given non M-L experiment $\mathcal{E}$ the pairwise least (most) informative among the types of M-L experiments which are pairwise at least (most) as informative as $\mathcal{E}$. We shall not attempt to discuss this (generally open) problem here and instead turn to the problem of characterizing important particular classes of M-L experiments in terms of the associated N-P functions.

Consider e.g. the M-L experiments $\mathcal{E} = (P_\theta : \theta \in \Theta)$ which are pairwise stationary in the sense that the informational content of a dichotomy $(P_{\theta_1}, P_{\theta_2})$ depends on the difference $\theta_2 - \theta_1$ only. If $\mathcal{E}$ possesses this property and if a is any number such that $a + \Theta \subseteqq \Theta$ then the M-L experiments $\mathcal{E}$ and $(P_{a+\theta} : \theta \in \Theta)$ are pairwise equivalent and are thus equivalent.

As we in the subsequent discussion of stationarity shall assume that Θ is the real line we do not need to distinguish between pairwise stationarity and stationarity.

Any translation experiment on the real line possesses this stationarity property. Furthermore strongly unimodal translation families are M-L experiments. It follows that the these distributions are dominated by the Lebesgue measure. In terms of the semigroup $(\gamma_h : h \geqq 0)$ this expresses that

$$\|\gamma_h - \gamma_k\| \leqq \|\gamma_{|k-h|} - \gamma_0\| \to 0 \text{ as } |h-k| \to 0.$$

Define a function F from R to [0,1] by choosing a number $\alpha_0 \in]0,1[$ and then putting:

(i) $F(x) = 1 - \gamma_{-x}(\alpha_0)$ when $x \leqq 0$.
(ii) $F(x) = 1 - y$ where $\gamma_x(y) = \alpha_0$ when $x \geqq 0$ and $\gamma_x(0) \leqq \alpha_0$.
(iii) $F(x) = 1$ otherwise; i.e. when $x \geqq 0$ and $\gamma_x(\alpha) > \alpha_0$.

It is then readily checked that F is continuous and monotonically increasing.

Note that when $k > 0$ then:

$$\sup_h \gamma_h(\alpha_0) = \lim_{h\to\infty} \gamma_h(\alpha_0) = \sup_h \gamma_k(\gamma_h(\alpha_0)) = \gamma_k(\sup \gamma_h(\alpha_0))$$

so that $\sup_h \gamma_h(\alpha_0) = 1$. Hence $\lim_{x\to-\infty} F(x) = 0$. If $F(x) \leqq \tau < 1$ for all x then, for $x > 0, \gamma_x(0) \leqq \alpha_0$ and $\alpha_0 = \gamma_x(1 - F(x)) \geqq \gamma_x(1-\tau)$ contradicting the fact, established above, that $\gamma_x(1-\tau) \to 1$ as $x \to \infty$. Thus F is a probability distribution function.

It may be checked that $1 - F(x-h) = \gamma_h(1 - F(x))$ when $h \geqq 0$ and $F(x) < 1$. [In fact if $x \leqq 0$ then $1 - F(x-h) = \gamma_{h-x}(\alpha_0)$ while $\gamma_h(1 - F(x)) = \gamma_h(\gamma_{-x}(\alpha_0)) = \gamma_{h-x}(\alpha_0)$. If $x > 0$ and $F(x) < 1$ then $\gamma_x(1 - F(x)) = \alpha_0$ i.e. $\gamma_{x-h}(\gamma_h(1 - F(x-h)) = \alpha_0$. If also $0 \leqq h \leqq x$ then $\gamma_{x-h}(1 - F(x-h)) = \alpha_0$ so that $1 - F(x-h) = \gamma_h(1 - F(x))$. If $x < h$ then $x - h < 0$ so that $1 - F(x-h) = \gamma_{h-x}(\alpha_0) = \gamma_{h-x}(\gamma_x(1 - F(x)) = \gamma_h(1 - F(x))$.]

If $F(a) < 1$ and $h \geqq 0$ this, in turn, implies that:

$$\int_{-\infty}^{a} \gamma_h'(1-F)dF = \int_{F \leqq a} \gamma_h'(1-F)dF - \int_{F=a} \gamma_h'(1-F)dF$$

$$= \int_0^{F(a)} \gamma_h'(1-z)dz = 1 - \gamma_h(1 - F(a)) = F(a-h) = F_h(a)$$

where F_θ, for any θ, denotes the θ-th translate of θ i.e. $F_\theta(x) \equiv_x F(x - \theta)$.

It follows that a F_θ maximal version of dF_θ/dF is given by:

$$dF_\theta/dF]_x = \gamma_\theta'(1 - F(x)) \text{ or } = \infty \text{ as} F(x) < 1 \text{ or } F(x) = 1.$$

Thus $F_\theta : \theta \in R$ has the M-L property and

$$\beta(0|F_0, F_\theta) = 1 - F(F^{-1}(1-\alpha)) = \gamma_\theta(\alpha)$$

Let $\mathcal{E} = (P_\theta : \theta \in \Theta)$ be any experiment having a finite parameterset Θ. If $a_\theta : \theta \in \Theta$ are real numbers then

$$\|\sum a_\theta P_\theta\| = 2[\sup \sum a_\theta \int \delta dP_\theta] - \sum_\theta a_\theta$$

where sup is taken over all testfunctions δ. Thus $\frac{1}{2}[\|\sum_\theta a_\theta P_\theta\| + \sum_\theta a_\theta]$ is the support function of the set of all powerfunctions of tests.

The following result provides therefor a link between the set of all powerfunctions of tests on the one hand and the behaviour of N-P functions on the other.

Let the total variation on a set A of a measure defined by a distribution function f be denoted as $\|f : A\|$. Then we may state:

Theorem 4.19 (Total variation norm for linear combinations).

Let $\mathcal{E} = (P_\theta : \theta \in \Theta)$ be a M-L experiment and consider parameter values $\theta_0, \theta_1, \cdots, \theta_m$ such that $\theta_0 \leqq \theta_i; i = 1, \cdots, m$.

Then for any numbers $a_1, \cdots, a_m$ the norm $\|\sum_{k=1}^m a_k P_{\theta_k}\|$ may be decomposed as:

$$\|\sum_{k=1}^m a_k P_{\theta_k}\| = \|\sum a_k \beta(\cdot|P_{\theta_0}, P_{\theta_k}) :]0,1]\| + r$$

where the first term on the right also equals the total variation of the P_{θ_0} absolutely continuous part of $\sum_{k=1}^m a_k P_{\theta_k}$ while r (consequently) is the total variation of the P_{θ_0} singular part of $\sum_{k=1}^m a_k P_{\theta_k}$. In particular

$$\|\sum_{k=1}^m a_k P_{\theta_k}\| = \|\sum_{k=1}^m a_k \beta(\cdot|P_{\theta_0}, P_{\theta_k})\|$$

when P_{θ_0} dominates $P_{\theta_1}, \cdots, P_{\theta_m}$.

In any case $r \leqq \sum_{k=1}^m |a_k| \beta(0|P_{\theta_0}, P_{\theta_k})$.

If $\theta_0 \leqq \theta_1 \leqq \cdots \leqq \theta_m$ then

$$r = \sum_{i=1}^{m-1} \|\sum_{k=i+1}^m a_k \beta(\cdot|P_{\theta_{i+1}}, P_{\theta_k}) :]0, \beta(0|P_{\theta_i}, P_{\theta_{i+1}})\|.$$

Proof:

Hence

$$r \leqq \sum_{i=1}^{m-1} \sum_{k=i+1}^{m} |a_k|[\beta_{i,k}(0) - \beta_{i+1,k}(0)]$$

$$= \sum_{i=1}^{m-1} \sum_{k=2}^{m} |a_k|[\beta_{i,k}(0) - \beta_{i+1,k}(0)]$$

where we put $\beta_{i,j}(0) = 0$ when $i \geqq j$. Interchanging $\sum_i$ and $\sum_k$ the last expression becomes $\sum_{k=2}^{m} |a_k|\beta_{1,k}(0) = \sum_{k=1}^{m} |a_k|\beta(0|P_{\theta_1}, P_{\theta_k})$.

□

The second limiting relation is proved by similar, but more involved, arguments.

Here are the details: Let $0 \leqq \alpha \leqq 1$. To each $h > 0$ there is then a testfunction δ_h so that $E_{\theta_0 - h}\delta_h = \alpha$ and $E_{\theta_0}\delta_h = \beta(\alpha|\theta_0 - h, \theta_0)$. Then $\int \delta_h dP^\bullet_{\theta_0} \leqq \kappa(\beta(\alpha|\theta_0 - h, \theta_0))$ so that $\limsup_{h \to 0} \int \delta_h dP^\bullet_{\theta_0} \leqq \kappa(\alpha)$.

It follows that

$$\begin{aligned}
&\limsup_{h \to 0} [\beta(\alpha|\theta_0 - h, \theta_0) - \alpha]/h \\
&= \limsup_{h \to 0} [E_{\theta_0}\delta_h - E_{\theta_0 - h}\delta_h]/h \leqq \limsup_{h \to 0} \int \delta_h dP^\bullet_{\theta_0} \leqq \kappa(\alpha).
\end{aligned}$$

Thus, since $\kappa(0) = \kappa(1) = 0$, convergence holds for $\alpha = 0$ and $\alpha = 1$. Assume so that $0 < \alpha < 1$.

It remains to show that $\liminf_{h \to 0} [\beta(\alpha|\theta_0 - h, \theta_0) - \alpha]/h \geqq \kappa(\alpha)$. If this was not so then there is a number $A < \kappa(\alpha)$ and a sequence $0 < h_n \downarrow 0$ so that

$$[\beta(\alpha|\theta_0 - h_n, \theta_0) - \alpha]/h_n \leqq A; n = 1, 2, \cdots$$

Restrict attention to numbers n such that $(A + \frac{1}{n})h_n \leqq 1 - \alpha$. (This is so for n sufficiently large). Suppress the subscript n, put $\epsilon = (A + \frac{1}{n})h/(1 - \alpha)$ and let δ be a slope optimal level α test-function so that $E_{\theta_0}\delta = \alpha$ while $\int \delta dP^\bullet_{\theta_0} = \kappa(\alpha)$. Consider the testfunction $\delta_\epsilon = (1 - \epsilon)\delta + \epsilon$. The above assumptions imply that

$$E_{\theta_0}\delta_\epsilon = (1 - \epsilon)\alpha + \epsilon = \alpha + \epsilon(1 - \alpha) = \alpha + (A + \frac{1}{n})h > \beta(\alpha|\theta_0 - h, \theta_0).$$

Hence, by optimality and since $\epsilon \to 0$ as $n \to \infty$:

$$\begin{aligned}
\alpha < E_{\theta_0 - h}\delta_\epsilon &= E_{\theta_0}\delta_\epsilon - h \int \delta_\epsilon dP^\bullet_{\theta_0} + ho(1) \\
&= \alpha + (A + \frac{1}{n})h - h \int \delta dP^\bullet_{\theta_0} - h\epsilon \int (1 - \delta) dP^\bullet_{\theta_0} + ho(1) \\
&= \alpha + (A + \frac{1}{n})h - h\kappa(\alpha) + ho(1)
\end{aligned}$$

so that $A + \frac{1}{n} > \kappa(\alpha) + o(1)$ yielding, as $n \to \infty$, the contradiction $\kappa(\alpha) > A \geqq \kappa(\alpha)$.

Assume so that $\mathcal{E}$ has the M-L property and that

$$\kappa(\alpha) = \lim[\beta(\alpha|\theta_0, \theta_0 + h) - \alpha]/h = \lim[\beta(\alpha|\theta_0 - h, \theta_0) - \alpha]/h$$

where lim is for $h \downarrow 0$ and where κ is a N-P slope function. If $h, k \geqq 0$ then, by theorem 4.19, $\|(P_{\theta_0 + h} - P_{\theta_0})/h - (P_{\theta_0 + k} - P_{\theta_0})/k\| \leqq$ the total variation of the function

$$\alpha \to \frac{1}{h}[\beta(\alpha|\theta_0, \theta_0 + h) - \alpha] - \frac{1}{k}[\beta(\alpha|\theta_0, \theta_0 + k) - \alpha]$$

Let $\alpha \in]0,1[$ and consider the fraction $[\beta(\alpha|\theta_0, \theta_0+h)-\alpha]/h$ where $h > 0$. Choose $\pi \in \prod_{\mathcal{E}}$ such that $\pi(\theta_0) = \alpha$. Putting $\kappa(\alpha) = \pi^{\bullet}(\theta_0)$ we find that

$$[\beta(\alpha|\theta_0, \theta_0 + h) - \alpha]/h = [\pi(\theta_0 + h) - \pi(\theta_0)]/h \to \kappa(\alpha) \text{ as } h \downarrow 0.$$

Extending the definition of κ to [0,1] by putting $\kappa(0) = \kappa(1) = 0$ we conclude that κ is continuous at $\alpha = 0$ and at $\alpha = 1$ and thus is a N-P slope function.

Trivially $[\beta(1|\theta_0, \theta_0 + h) - 1]/h = 0 \to \kappa(1)$. If $0 < \alpha$ then, as we have seen, $\kappa(\alpha) = \lim[\beta(\alpha|\theta_0, \theta_0 + h) - \alpha]/h$. Thus

$$\begin{aligned}\kappa(\alpha) &\geqq \limsup_{h\to 0} \frac{(1-\beta(0|\theta_0, \theta_0+h))\alpha + \beta(0|\theta_0, \theta_0+h) - \alpha}{h} \\ &= (1-\alpha)\limsup_{h\to 0} \beta(0|\theta_0, \theta_0+h)/h.\end{aligned}$$

Letting $\alpha \to 0$ we find that $\limsup_{h\to 0} \beta(0|\theta_0, \theta_0 + h)/h = 0$ so that

$$[\beta(\alpha|\theta_0, \theta_0 + h) - \alpha]/h \to \kappa(\alpha) \text{ for all } \alpha \in [0,1].$$

In order to apply the theorem we need also to consider the left difference fraction $[\beta(\alpha|\theta_0 - h, \theta_0) - \alpha]/h$. By the M-L property the function $\beta(\cdot|\theta_0 - h, \theta_0)$ decreases pointwise to a N-P function γ as $h \downarrow 0$. By Dini's Lemma this convergence is uniform on [0,1]. The convergence is however also in the sense of total variation.

Choose $\alpha \in]0,1[$ and put $\alpha_h = \pi(\theta_0 - h)$ where $\pi \in \prod_{\mathcal{E}}$ and $\pi(\alpha_0) = \alpha$. Then $\beta(\alpha_h|\theta_0 - h, \theta_0) - \gamma(\alpha_h) \to 0$ while $\alpha_h \to \pi(\theta_0) = \alpha$ since any π is continuous at θ_0. Thus $\gamma(\alpha_h) \to \gamma(\alpha)$ so that $\beta(\alpha_h|\theta_0 - h, \theta_0) \to \gamma(\alpha)$. On the other hand:

$$\frac{\beta(\alpha_h|\theta_0 - h, \theta_0) - \alpha_h}{h} = \frac{\pi(\theta_0) - \pi(\theta_0 - h)}{h} \to \pi^{\bullet}(\theta_0) = \kappa(\alpha).$$

Thus $\beta(\alpha_h|\theta_0 - h, \theta_0) \to \alpha$ so that $\gamma(\alpha) = \alpha$. Hence $\gamma(\alpha) \equiv_\alpha \alpha$ so that $\beta(\alpha|\theta_0 - h, \theta_0) \downarrow \alpha$ uniformly in α as $h \downarrow 0$.

Consider $\alpha_h = \omega_h(\alpha)$ as a function of $\alpha \in]0,1[$ for each $h > 0$. Then $\beta(\omega_h(\alpha)|\theta_0 - h, \theta_0) = \alpha; 0 < \alpha < 1$. It follows that if $\alpha > \beta(0|\theta_0 - h, \theta_0)$, which is the case when h is sufficiently small, then $x = \omega_h(\alpha)$ is the unique solution of the equation $\beta(x|\theta_0 - h, \theta_0) = x$. If $0 < \alpha \leqq \beta(0|\theta_0 - h, \theta - 0)$ then $\omega_h(\alpha) = \pi(\theta_0 - h) = 0$. Put also $\omega_h(0) = 0$ and $\omega_h(1) = \omega_h(1-)$. Then ω_h is a convex function (actually a Lorenz function). If $\alpha \in]0,1[$ then again:

$$[\alpha - \omega_h(\alpha)]/h = [\pi(\theta_0) - \pi(\theta_0 - h)]/h \to \pi^{\bullet}(\theta_0) = \kappa(\alpha)$$

and trivially this holds also for $\alpha = 0$. The number $\omega_h(1) = \omega_h(1-)$ is the smallest number t_h such that $\beta(t_h|\theta_0 - h, \theta_0) = 1$. There is then a powerfunction $\pi \in \prod_{\mathcal{E}}$ such that $\pi(\theta_0) = 1$ while $\pi(\theta_0 - h) = t_h$. Then:

$$A = \| -\frac{1}{h}\beta(\cdot|\theta_0,\theta_0) + \frac{1}{h}\beta(\cdot|\theta_0,\theta_0+h) + \frac{1}{h}\beta(\cdot|\theta_0,\theta_1) - \frac{1}{h}\beta(\cdot|\theta_0,\theta_1+h) :]0,1]\|,$$
$$B_1 = \|\frac{1}{h}\beta(\cdot|\theta_0+h,\theta_0+h) + \frac{1}{h}\beta(\cdot|\theta_0+h,\theta_1) - \frac{1}{h}\beta(\cdot|\theta_0+h,\theta_1+h) :]0,\beta(0|\theta_0,\theta_0+h)]\|,$$
$$B_2 = \|\frac{1}{h}\beta(\cdot|\theta_1,\theta_1) - \frac{1}{h}\beta(\cdot|\theta_1,\theta_1+h) :]0,\beta(0|\theta_0+h,\theta_1)]\| \text{ and}$$
$$B_3 = \|\frac{1}{h}\beta(\cdot|\theta_1+h,\theta_1+h) :]0,\beta(0|\theta_1,\theta_1+h)]\|.$$

Clearly

$$\|(P_{\theta_0+h} - P_{\theta_0})/h - (P_{\theta_1+h} - P_{\theta_1})/h\| \to \|P^\bullet_{\theta_0} - P^\bullet_{\theta_1}\| \text{ as } h \downarrow 0.$$

The proof is now established by showing, as $h \downarrow 0$, that:

$$A \to \|\kappa(\cdot|\theta_0) - \kappa(\beta(\cdot|\theta_0,\theta_1) :]0,1]\|, B_1 \to 0,$$
$$\limsup B_2 \leqq \|\kappa(\cdot|\theta_1) :]0,\beta(0|\theta_0,\theta_1)]\| \text{ and } B_3 \to 0.$$

This is seen as follows:

A may be rewritten as

$$A = \|\frac{1}{h}[\beta(\cdot|\theta_0,\theta_0+h) - \lambda] - \frac{1}{h}[\beta(\beta(\cdot|\theta_0,\theta_1)|\theta_1,\theta_1+h) - \beta(\cdot|\theta_0,\theta_1)] :]0,1]\|$$
$$= \|\frac{1}{h}[]_1 - \frac{1}{h}[]_2 :]0,1]\|$$

where

$$\frac{1}{h}[]_1 \to \kappa(\cdot|\theta_0) \text{ and } \frac{1}{h}[]_2 \to \kappa(\beta(\cdot|\theta_0,\theta_1)|\theta_1)$$

in total variation on]0,1].

The expression for B_1 may be rewritten as:

$$B_1 = \|\frac{1}{h}\lambda - \frac{1}{h}[\beta(\beta(\cdot|\theta_0+h,\theta_1)|\theta_1,\theta_1+h) - \beta(\cdot|\theta_0+h,\theta_1)] :]0,\alpha_h]\|$$

where $\alpha_h = \beta(0|\theta_0,\theta_0+h) \downarrow 0$. Hence provided $\alpha_h < \alpha_0$:

$$B_1 \leqq \frac{1}{h}\|\lambda :]0,\alpha_h]\| + \|\kappa_{\theta_1}(\beta(\cdot|\theta_0+h,\theta_1) :]0,\alpha_0]\| + o(1).$$

It follows, since $\alpha_h(h) = [\beta(0|\theta_0,\theta_0+h) - 0]/h \to \kappa(0|\theta_0) = 0$, that $\limsup B_1 \leqq \|\kappa_{\theta_1} :]0,\alpha_0]\| \downarrow 0$ as $\alpha_0 \downarrow 0$.

Assume so that $\mathcal{E}$ is a M-L experiment such that $\kappa(\alpha|\theta)$ is continuous in θ.

By corollary 5.3

$\|P^{\bullet}_{\theta_0} - P^{\bullet}_{\theta_1}\| \leqq \|\kappa(\cdot|\theta_0) - \kappa(\beta(\cdot|\theta_0,\theta_1)|\theta_1) :]0,1]\| + \|\kappa(\cdot|\theta_1) :]0,\beta(0|\theta_1,\theta_1)]\|$ when $\theta_0 < \theta_1$.

If $\theta_1 \downarrow \theta_0$ or if $\theta_0 \uparrow \theta_1$ then $\|\kappa(\cdot|\theta_1) - \kappa(\cdot|\theta_0)\| \to 0$. In the first case

$$\begin{aligned}\|P^{\bullet}_{\theta_0} - P^{\bullet}_{\theta_1}\| &\leqq \|\kappa(\cdot|\theta_0) - \kappa(\beta(\cdot|\theta_0,\theta)|\theta_0) :]0,1]\| \\ &+ \|\kappa(\beta(\cdot|\theta_0,\theta_1)|\theta_0) - \kappa(\beta(\cdot|\theta_0,\theta_1)|\theta_1) :]0,1]\| \\ &+ \|\kappa(\cdot|\theta_0) :]0,\beta(0|\theta_0,\theta_1)]\| + \|\kappa(\cdot|\theta_1) - \kappa(\cdot|\theta_0) :]0,1]\| \\ &\leqq o(1) + 2\|\kappa(\cdot|\theta_0) - \kappa(\cdot|\theta_1)\| \to 0\end{aligned}$$

In the second case

$$\begin{aligned}\|P^{\bullet}_{\theta_0} - P^{\bullet}_{\theta_1}\| &\leqq \|\kappa(\cdot|\theta_0) - \kappa(\cdot|\theta_1)\| \\ &+ \|\kappa(\cdot|\theta_1) - \kappa(\beta(\cdot|\theta_0,\theta_1) :]0,1]\| + \|\kappa(\cdot|\theta_1) :]0,\beta(0|\theta_0,\theta_1)]\| \to 0.\end{aligned}$$

□

In terms of powerfunctions continuous differentiability of M-L experiments may be expressed as follows:

Corollary 5.5 (Continuous derivatives of powerfunctions).

Let $\mathcal{E} = (P_\theta : \theta \in \Theta)$ be an everywhere differentiable M-L experiment. Then the derivative $P^{\bullet}_\theta$ is continuous in θ for the total variation distance if and only if the family $(\pi^{\bullet} : \pi \in \prod_{\mathcal{E}})$ of differentiated powerfunctions is equicontinuous on Θ.

Proof:
If $P^{\bullet}_\theta$ is continuous in θ then, as we have seen, $\kappa(\alpha|\theta)$ is continuous in θ for each α. This implies for these slope functions that $\kappa(\alpha|\theta)$ is jointly continuous in (α,θ). Combining this with the fact that $\pi^{\bullet} = \kappa(\pi|\cdot)$ when $\pi \in \prod_{\mathcal{E}}$ we find that $\prod_{\mathcal{E}}$ as well as $\{\pi^{\bullet} : \pi \in \prod_{\mathcal{E}}\}$ is equicontinuous. (In fact $\prod_{\mathcal{E}}$ is uniformly Lipschitz on compacts).

Assume so that the family $(\pi^{\bullet} : \pi \in \prod_{\mathcal{E}})$ is uniformly equicontinuous. It suffices, by the theorem to show that $\kappa(\alpha|\theta)$ is continuous in θ for a given $\alpha \in]0,1[$. Consider then numbers $\theta_n; n = 1,2\cdots$ converging to θ. Let $\pi,\pi_1,\pi_2,\cdots$ be the powerfunctions determined by the requirements $\alpha = \pi(\theta) = \pi_1(\theta_1) = \pi_2(\theta_2) = \cdots$. Then $\kappa(\alpha|\theta_n) = \kappa(\pi_n(\theta_n)|\theta_n) = \pi^{\bullet}_n(\theta_n)$ while $\kappa(\alpha|\theta) = \kappa(\pi(\theta)|\theta) = \pi^{\bullet}(\theta)$.

It suffices therefor to show that $\pi^{\bullet}_n(\theta_n) \to \pi^{\bullet}(\theta)$. Now

$$\pi^{\bullet}_n(\theta_n) - \pi^{\bullet}(\theta) = [\pi^{\bullet}_n(\theta_n) - \pi^{\bullet}_n(\theta)] + [\pi^{\bullet}_n(\theta) - \pi^{\bullet}(\theta)]$$

continuous family $(\tilde{\kappa}(\cdot|\theta) : \theta \in \Theta)$ of slope functions and if $\kappa(\cdot|\theta) \leqq \tilde{\kappa}(\cdot|\theta)$ for all $\theta \in \Theta$ then, letting $\tilde{y}$ denote the corresponding solution for $\tilde{\kappa}, \tilde{y}(\theta|\theta_0, \alpha_0) \leqq y(\theta|\theta_0, \alpha_0)$ or $\geqq y(\theta|\theta_0, \alpha_0)$ as $\theta \leqq \theta_0$ or $\theta \geqq \theta_0$.

Furthermore if w is any function on Θ such that $w(\theta_0) = \alpha_0$ and having a derivative $w^\bullet$ such that $w^\bullet(\theta) \leqq \kappa(w(\theta)|\theta)$ for all θ then $w(\theta) \geqq y(\theta|\theta_0, \alpha_0)$ or $\leqq y(\theta|\theta_0, \alpha_0)$ as $\theta \leqq \theta_0$ or $\theta \geqq \theta_0$.

Finally if v is any function on Θ such that $v(\theta_0) = \alpha$ and having a derivative $v^\bullet$ such that $v^\bullet(\theta) \geqq \kappa(v(\theta)|\theta)$ for all θ where the inequality is strict whenever $\kappa(\alpha|\theta) \equiv_\alpha 0$, then $v(\theta) \leqq y(\theta|\theta_0, \alpha_0)$ or $\geqq y(\theta|\theta_0, \alpha_0)$ as $\theta \leqq \theta_0$ or $\theta \geqq \theta_0$.

Proof:

The local existence theory for 1-st order differential equations implies that there is a $a > 0$ such that there is a function $y = y(\cdot|\theta_0, \alpha_0)$ from $[\theta_0 - a, \theta_0 + a]$ to $]0,1[$ such that $y^\bullet(\theta) = \kappa(y(\theta)|\theta)$ when $\theta \in [\theta_0 - a, \theta_0 + a]$ while $y(\theta_0) = \alpha_0$. The uniqueness part of this theory implies that there is at most one such function from Θ to $[0,1]$. The desired global solution may be obtained as a maximal extension of the fragment on $[\theta_0 - a, \theta_0 + a]$.

Consider so two continuous families $\kappa_1(\cdot|\theta) : \theta \in \Theta$ and $\kappa_2(\cdot|\theta) : \theta \in \Theta$ such that $\kappa_1(\cdot|\theta) \leqq \kappa_2(\cdot|\theta)$ for all θ. Put $\kappa_2^\epsilon(\alpha|\theta) = \kappa_2(\alpha|\theta) + \epsilon \min\{\alpha, 1 - \alpha\}$ when $\epsilon > 0$. Let z_ϵ be the solution for κ_2^ϵ, let z be the solution for κ_2 and let y be the solution for κ_1. Thus $z_\epsilon(\theta_0) = z(\theta_0) = y(\theta_0) = \alpha_0$ while for any $\theta \in \Theta : z_\epsilon^\bullet(\theta) = \kappa_2^\epsilon(z_\epsilon(\theta)|\theta), z^\bullet(\theta) = \kappa_2(z(\theta)|\theta)$ and $y^\bullet(\theta) = \kappa_1(y(\theta)|\theta)$.

Then $z_\epsilon(\theta) \geqq z(\theta)$ when $\theta \geqq \theta_0$ while $z_\epsilon(\theta) \leqq z(\theta)$ when $\theta \leqq \theta_0$. If $\theta > \theta_0$ then this may be argued as follows: Firstly $z_\epsilon(\theta_0) = z(\theta_0) = \alpha$ while $z_\epsilon^\bullet(\theta_0) > z^\bullet(\theta_0)$. Thus $z_\epsilon(\theta) > z(\theta)$ when $\theta \geqq \theta_0$ and θ is sufficiently close to θ_0. If $z_\epsilon(\theta) < z(\theta)$ for some $\theta > \theta_0$ then there is a $\theta_1 > \theta_0$ such that $z_\epsilon(\theta_1) = z(\theta_1)$ while $z_\epsilon(\theta) > z(\theta)$ when $\theta \in]\theta_0, \theta_1[$. Then $z_\epsilon^\bullet(\theta_1) \leqq z^\bullet(\theta_1)$ which is impossible since $z_\epsilon^\bullet(\theta_1) = \kappa_2^\epsilon(z_\epsilon(\theta_1)|\theta_1) = \kappa_2^\epsilon(z(\theta_1)|\theta_1) > \kappa_2(z(\theta_1)|\theta_1) = z^\bullet(\theta_1)$. If $\theta < \theta_0$ then the derived inequality may likewise be argued by comparing slopes at points of intersection of the graphs of z and of z_ϵ.

By about the same kind of arguments we find also that $z_\epsilon(\theta) \geqq y(\theta)$ when $\theta \geqq \theta_0$ while $z_\epsilon(\theta) \leqq y(\theta)$ when $\theta \leqq \theta_0$. Furthermore $z_\epsilon(\theta)$ increases monotonically in ϵ when $\theta \geqq \theta_0$ while $z_\epsilon(\theta)$ decreases monotonically when $\theta \leqq \theta_0$. It is readily checked that $\theta \rightarrow \lim_{\epsilon \rightarrow 0} z_\epsilon(\theta)$ satisfies the same differential equation as z and trivially $\lim_{\epsilon \rightarrow 0} z_\epsilon(\theta_0) = \alpha_0$. Hence $\lim_{\epsilon \rightarrow 0} z_\epsilon(\theta) = z(\theta); \theta \in \Theta$ so that $z(\theta) \geqq y(\theta)$ or $\leqq y(\theta)$ as $\theta \geqq \theta_0$ or $\theta \leqq \theta_0$.

Consider next any function w from Θ to $[0,1]$ such that $w(\theta_0) = \alpha_0$ while $w^\bullet(\theta) \leqq \kappa(w(\theta)|\theta)$ for all θ. Put $\kappa^\epsilon(\alpha|\theta) = \kappa(\alpha|\theta) + \epsilon \min\{\alpha, 1 - \alpha\}$ for $\epsilon > 0$. (κ^ϵ is introduced in order to be able to work with a strictly positive slope function on]0,1[). Let y_ϵ be the unique function such that $y_\epsilon^\bullet(\theta) = \kappa^\epsilon(y(\theta)|\theta); \theta \in \Theta$ while $y_\epsilon(\theta_0) = \alpha_0$.

Let α_0, α_1 and τ be numbers in]0,1[and put

$$z(\theta) = (1-\tau)y(\theta|\theta_0, \alpha_0) + \tau y(\theta|\theta_0, \alpha_1).$$

Then, since $\kappa(\cdot|\theta)$ is concave,

$$\begin{aligned} z^{\bullet}(\theta) &= (1-\tau)y^{\bullet}(\theta|\theta_0, \alpha_0) + \tau y^{\bullet}(\theta|\theta_0, \alpha_1) \\ &= (1-\tau)\kappa(y(\theta|\theta_0, \alpha_0)|\theta) + \tau\kappa(y(\theta|\theta_0, \alpha_1)|\theta) \\ &\leqq \kappa(z(\theta)|\theta). \end{aligned}$$

On the other hand $z(\theta_0) = (1-\tau)\alpha_0 + \tau\alpha_1$. Hence, by the theorem: $y(\theta|\theta_0, \alpha) \leqq z(\theta)$ or $\geqq z(\theta)$ as $\theta \leqq \theta_0$ or $\theta \geqq \theta_0$.

It follows that $y(\theta|\theta_0, \alpha)$ is convex or concave in α as $\theta \leqq \theta_0$ or $\theta \geqq \theta_0$. In particular $y(\theta_1)$ is a concave function of $y(\theta_0) > 0$ when y varies in $\prod$ provided $\theta_0 < \theta_1$.

This shows that $\prod$ satisfies both condition (a) and condition (b) of section 4. The closure $\overline{\prod}$ of $\prod$ for pointwise convergence satisfies then also (a) and (b). Indeed $\overline{\prod}$ consists of the functions in $\prod$ and the set of indicator functions which are pointwise limits of functions in $\prod$.

It follows from theorem 4.13 that $\overline{\prod} = \prod_{\mathcal{E}}$ for a M-L experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$. If $\pi \in \prod$ then $\kappa(\pi(\theta)|\theta, \mathcal{E}) = \pi^{\bullet}(\theta) = \kappa(\pi(\theta)|\theta)$ so that $\kappa(\cdot|\theta, \mathcal{E}) = \kappa(\cdot|\theta)$ when $\theta \in \Theta$. The continuity of $\kappa(\alpha|\theta)$ for given α implies by corollary 5.5 that $P_\theta^{\bullet}$ is continuous in θ for total variation distance. Finally uniqueness follows from the uniqueness of solutions of our differential equations.

□

In spite of the uniqueness part of our last theorem there is in general a multitude of continuously differentiable non M-L experiments having the same slope functions as $\mathcal{E}$. In fact if $\mathcal{F}$ is any continuously differentiable experiment then our results imply that there is a M-L experiment which is locally equivalent with $\mathcal{F}$. If $\mathcal{F}$ does not have the M-L property then $\mathcal{E}$ and $\mathcal{F}$ can not be equivalent.

If $\mathcal{E}$ is any experiment which is differentiable at $\theta = \theta_0$ and if $\kappa(\cdot|\theta_0) \not\equiv 0$ then, by Torgersen (1985), $\mathcal{E}$ is locally equivalent at θ_0 to a unique strongly unimodal translation experiment $\mathcal{G} = (G_\theta : \theta \in \Theta)$. The distribution function G along with all its translates constitute the total set of solutions of the differential equation

$$G'(G^{-1}(1-\alpha)) = \kappa(\alpha); 0 \leqq \alpha \leqq 1.$$

As a particular case consider the Cauchy translation experiment $\mathcal{E}$ determined by the Cauchy density $x \to \frac{1}{\pi(1+x^2)}$. As is well known this experiment is not a M-L experiment i.e. $x \to \log(1+x^2)$ is not convex. Since $\mathcal{E}$ is a translation experiment the slope function $\kappa(\cdot|\theta, \mathcal{E})$ does not depend on θ. Thus the constructed strongly unimodal translation experiment $\mathcal{G}$ is everywhere locally equivalent to $\mathcal{E}$. By our next theorem $\mathcal{G}$ is pairwise more

6. COMPARISON FOR GIVEN LOSS FUNCTIONS. APPLICATIONS TO SELECTION PROBLEMS.

Consider a decision problem given by an experiment $\mathcal{E} = (\chi, \mathcal{A}, P_\theta : \theta \in \Theta)$, a decision space $(T, \mathcal{D})$ (i.e. a measurable space) and a loss function L on T. Thus $L = (L_\theta : \theta \in \Theta)$ is a family of extended real valued measurable functions on $(T, \mathcal{D})$. We shall for simplicity assume that L_θ, for each θ, is bounded from below. This condition is assumed fulfilled for the decision problems considered in this paper.

If not otherwise stated the parameter set Θ is an arbitrary, but fixed set.

Within this set up a decision rule ρ in the traditional sense is a Markov kernel from the sample space of $\mathcal{E}$ to the decision space $(T, \mathcal{D})$.

The operational characteristic of ρ is the experiment $(P_\theta \rho : \theta \in \Theta)$ where $(P_\theta \rho)(D) = \int \rho(D|x) P_\theta(dx); D \in \mathcal{D}, \theta \in \Theta$.

The risk $r(\theta|\rho)$ incurred by using the decision rule ρ when θ prevails may be expressed in terms of the operational characteristic as $\int L_\theta(t) P_\theta \rho(dt)$ or as a double expectation $\int [\int L_\theta(t) \rho(dt|x)] P_\theta(dx)$. The risk function of the decision rules ρ is $r(\theta|\rho)$ as a function of θ.

We shall need the set Λ of prior distributions λ on Θ having finite supports. A distribution λ in Λ may then be identified with the weights $\lambda_\theta : \theta \in \Theta$ it assignes to the parameter points. We do not require that the prior distributions in Λ are probability distributions. Thus $(\lambda_\theta : \theta \in \Theta)$ are the weights of a prior distribution λ in Λ if and only if these weights are all non negative and if the set $\{\theta : \lambda_\theta > 0\}$ is finite.

The Bayes risk of ρ may now be written:

$$\sum_\theta \lambda_\theta r(\theta|\rho) = \int [\int [\sum_\theta \lambda_\theta(t) f_\theta] \rho(dt|\cdot)] d\mu$$

where μ is a non negative measure such that P_θ is μ-absolutely continuous with density $f_\theta = dP_\theta / d\mu$ when $\lambda_\theta > 0$. The minimum Bayes risk will be denoted by $b(\lambda|L, \mathcal{E})$ or just by $b(\lambda|L)$. Thus $b(\lambda|L, \mathcal{E}) = \int [\inf_t (\sum_\theta \lambda_\theta L_\theta(t) f_\theta)] d\mu$. The minimum Bayes risk $b(\lambda|L, \mathcal{E})$ as a function of the variables λ, L and $\mathcal{E}$ plays an important role in decision theory in general and in the theory of comparison of experiments in particular.

Considering the loss function L and the experiment $\mathcal{E}$ as fixed the fundamental observation is that $b(\cdot|L, \mathcal{E})$ provides, by the very definition, the lower support function of the lower boundary of the set of risk functions.

Trivially $\sum_\theta s(\theta) \lambda_\theta \geqq b(\lambda|L, \mathcal{E})$ for any function s on Θ such that $s(\theta) \geqq r(\theta|\rho)$ for all θ for some decision rule ρ.

One benefit of admitting generalized decision rules is that the set of generalized decision rules is compact for the pointwise topology on $\mathcal{L}(\mathcal{E}) \times M(T, \mathcal{D})$. We shall from here on permit ourselves to work freely with generalized decision rules without always referring to them as generalized. It should however be noted that the fundamental quantity

$$b(\lambda|L, \mathcal{E}) = \inf_{\rho} \sum_{\theta} \lambda_\theta (P_\theta \rho L_\theta)$$

is unaltered by the admittance of generalized decision rules.

Another useful fact is the following: Say that a traditional decision rule ρ is finitely supported if $\rho(D|\cdot) = 1$ everywhere for some finite union D of $\mathcal{D}$ atoms. Then the set of finitely supported traditional decision rules (i.e. Markov kernels) is dense within the set of all decision rules for the topology of pointwise convergence on $L(\mathcal{E}) \times M(T, \mathcal{D})$.

We are now in a position to present the fundamental characterization of the lower boundary of the risk set due to LeCam.

Theorem 6.1 (Support function description of the lower boundary of the risk set).

Let the decision problem $\mathcal{E}, (T, \mathcal{D})$ and L be as above. Then the following two conditions are equivalent for an extended real valued function s on Θ:

(i) There is a decision rule ρ such that $r(\theta|\rho) \leqq s(\theta)$ for all $\theta \in \Theta$.
(ii) $b(\lambda|L, \mathcal{E}) \leqq \sum_{\theta} \lambda_\theta s(\theta)$ for all $\lambda \in \Lambda$.

Proof:
Suppose $b(\lambda|L, \mathcal{E}) \leqq \sum_{\theta} \lambda_\theta s(\theta)$ for all $\lambda \in \Lambda$. We must prove the existence of a decision rule ρ such that $P_\theta \rho L_\theta \leqq s(\theta); \theta \in \Theta$.

If $s(\theta) = \infty$ for some θ then the last inequality is trivial. It follows that we without loss of generality may assume that $s(\theta) < \infty$ for all $\theta \in \Theta$.

Consider the 2-person null-sum game with pay off function $(\lambda, \rho) \to \sum_{\theta} \lambda_\theta P_\theta \rho L_\theta - \sum_{\theta} \lambda_\theta s(\theta)$. This pay off function is affine in λ as well as in ρ. If L is bounded then it is continuous in ρ. In general it is at least lower semicontinuous in ρ. Furthermore the set of strategies for player II is compact for the pointwise convergence on $L(\mathcal{E}) \times M(T, \mathcal{D})$.

It follows by standard minimax theory that this game has a value and that player II has a minimax strategy ρ_0. The assumptions imply that the lower value is non positive so that

$$\sum_{\theta} \lambda_\theta P_\theta \rho_0 L_\theta \leqq \sum_{\theta} \lambda_\theta s(\theta)$$

for all $\lambda \in \Lambda$. Inserting the one point distribution for λ we find that

$$P_\theta \rho_0 L_\theta \leqq s(\theta); \theta \in \Theta.$$

In particular $(\sum_{i=1}^{r} p_i\mathcal{E}_i, L) \geqq (\sum_{i=1}^{r} p_i\mathcal{F}_i, W)$ when $(\mathcal{E}_i, L) \geqq (\mathcal{F}_i, W); i = 1, \cdots, r$.

Proof:
This follows immediately from the fact that $b(\lambda|L, \mathcal{E})$ is affine in $\mathcal{E}$.

□

The minimum Bayes risk for product experiments may conveniently be expressed in terms of updated prior distributions as follows:

Suppose the experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ and $\mathcal{F} = (Q_\theta : \theta \in \Theta)$ are realized by observing, respectively, the independent variables X and Y. Then the minimum Bayes risk $b(\lambda|L, \mathcal{E} \times \mathcal{F})$ may be expressed as

$$b(\lambda|L, \mathcal{E} \times \mathcal{F}) = Eb(\lambda(\cdot|Y)|L, \mathcal{E}) = Eb(\lambda(\cdot|X)|L, \mathcal{F})$$

where $\lambda(\cdot|Y)$ and $\lambda(\cdot|X)$ denote, respectively, the a posteriori distributions of θ given Y, respectively, the a posteriori distribution of θ given X.

Consider also two other independent variables $\tilde{X}$ and $\tilde{Y}$ realizing, respectively, experiments $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{F}}$. Assume that the decision problem $(\mathcal{E}, L)$ is κ_1 deficient w.r.t. the decision problem $(\tilde{\mathcal{E}}, L)$, and that the decision problem $(\mathcal{F}, L)$ is κ_2 deficient w.r.t. the decision problem $(\tilde{\mathcal{F}}, L)$. Then:

$$\begin{aligned} b(\lambda|L, \mathcal{E} \times \mathcal{F}) &\leqq Eb(\lambda(\cdot|Y)|L, \tilde{\mathcal{E}}) + \sum_\theta \lambda_\theta \kappa_2(\theta) \\ &= b(\lambda|L, \tilde{\mathcal{E}} \times \mathcal{F}) = Eb(\lambda(\cdot|\tilde{X})L, \mathcal{F}) + \sum_\theta \lambda_\theta \kappa_2(\theta) \\ &\leqq Eb(\lambda(\cdot|\tilde{X})|L, \tilde{\mathcal{F}}) + \sum_\theta \lambda_\theta \kappa_1(\theta) + \sum_\theta \lambda_\theta \kappa_2(\theta) \\ &= b(\lambda|\tilde{\mathcal{E}} \times \tilde{\mathcal{F}}) + \sum_\theta \lambda_\theta[\kappa_1(\theta) + \kappa_2(\theta)]. \end{aligned}$$

This proves:

Corollary 6.4 (Comparison of products of decision problems).

Assume, for $i = 1, 2, \cdots, r$, that the decision problem $(\mathcal{E}_i, L)$ is κ_i deficient w.r.t. the decision problem $(\mathcal{F}_i, L)$.

Then the decision problem $(\prod_{i=1}^{r} \mathcal{E}_i, L)$ is $\sum_{i=1}^{r} \kappa_i$ deficient w.r.t the decision problem $(\prod_{i=1}^{r} \mathcal{F}_i, L)$.

If these integrals are finite and P_0 and P_1 are not mutually singular then this yields $P_0(S)/P_1(S) \geqq P_0(S)^{\frac{1}{n}}$. Letting $n \to \infty$ we find that $P_0(S) \geqq P_1(S)$. This proves:

Proposition 6.6

Let $\mathcal{E}^S$ be the experiment obtained from the experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ by selection on a non ancillary event S. Assume that $0 < \int_S (dP_{\theta_2}/dP_{\theta_1})^n dP_{\theta_1} < \infty$ for $n = 1, 2, \cdots$ and all $\theta_1, \theta_2 \in \Theta$. Then $\mathcal{E}^S$ is not pairwise at least as informative as $\mathcal{E}$.

Consider also a decision space $(T, \mathcal{D})$ equipped with a loss function $L = (L_\theta(t) : t \in T, \theta \in \Theta)$. We shall assume that there is a decision t in T such that $L_\theta(t) < \infty$ for all θ.

Let us also assume that $P_\theta(S) < 1$ for all θ such that $\mathcal{E}^{S^c}$ is well defined. We may thus compare the decision problems $(\mathcal{E}, L), (\mathcal{E}^S, L)$ and $(\mathcal{E}^{S^c}, L)$. One might expect that if information is generally increased by selection on S then information is generally lost by selection on S^c. We shall now see that this is so for several important notions of information. The following result state that this is so for any given loss function:

Theorem 6.7

Let the experiment $\mathcal{E}$, the loss function L and the selection set S be as above. Then $(\mathcal{E}^{S^c}, L) \leqq (\mathcal{E}, L)$ provided $(\mathcal{E}^S, L) \geqq (\mathcal{E}, L)$.

Proof:

We may without loss of generality assume that Θ is finite. Put $\mu = \sum_\theta P_\theta$ and $f_\theta = dP_\theta/d\mu$. The inequality $(\mathcal{E}^S, L) \geqq (\mathcal{E}, L)$ may be written:

$$\int_S [\bigwedge_t \sum_\theta \lambda_\theta (f_\theta/P_\theta(S)) L_\theta(t)] d\mu \leqq \int [\bigwedge_t \sum_\theta \lambda_\theta f_\theta L_\theta(t)] d\mu; \lambda \in \Lambda.$$

Replacing λ_θ with $\lambda_\theta P_\theta(S)$ this may also be written:

$$\int_S [\bigwedge_t \sum_\theta \lambda_\theta f_\theta L_\theta(t)] d\mu \leqq \int [\bigwedge_t \sum_\theta \lambda_\theta P_\theta(S) f_\theta L_\theta(t)] d\mu; \lambda \in \Lambda$$

or

$$\int_{S^c} [\bigwedge_t \sum_\theta \lambda_\theta f_\theta L_\theta(t)] d\mu \geqq \int \{[\bigwedge_t \sum_\theta \lambda_\theta f_\theta L_\theta(t)] - [\bigwedge_t \sum_\theta \lambda_\theta P_\theta(S) f_\theta L_\theta(t)]\} d\mu$$

$\geqq \int [\bigwedge_t \sum_\theta \lambda_\theta P_\theta(S^c) f_\theta L_\theta(t)] d\mu; \lambda \in \Lambda$ where the last $\geqq$ follows by super additivity.

Replacing λ_θ by $\lambda_\theta / P_\theta(S^c)$ we obtain the informational inequality $(\mathcal{E}, L) \geqq (\mathcal{E}^{S^c}, L)$. □

If $\mathcal{E}$ is at least as informative as $\mathcal{F}$ then $H(\cdot|\mathcal{E}) \leqq H(\cdot|\mathcal{F})$. If the latter inequality holds then we shall say that $\mathcal{E}$ *is at least as informative as* $\mathcal{F}$ *for the Hellinger ordering of experiments.*

If Pand Q are probability measures on the same measurable space then $\int \sqrt{dPdQ}$ is called *the affinity* between Pand Q. Considering an experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ and parameter points θ_1 and θ_2 we find that the affinity between P_{θ_1} and P_{θ_2} is $H(\lambda|\mathcal{E})$ where $\lambda_{\theta_1} = \lambda_{\theta_2} = \frac{1}{2}$ or $= 1$ as $\theta_1 \neq \theta_2$ or as $\theta_1 = \theta_2$.

Analogously with proposition 6.8 we have:

Proposition 6.10

Let $\{S_1, S_2, \cdots\}$ be a measurable partitioning of the sample space of $\mathcal{E} = (P_\theta : \theta \in \Theta)$ such that $P_\theta(S_i) > 0; \theta \in \Theta, i = 1, 2, \cdots$. Assume that $\mathcal{E}$ is regular in the sense that $\bigwedge_{\theta \in F} P_\theta \neq 0$ for any finite subset F of Θ.

Suppose $H(\cdot|\mathcal{E}^{S_i}) \leqq H(\cdot|\mathcal{E}); i = 1, 2, \cdots$. Then $\mathcal{E}^{S_i} \sim \mathcal{E}; i = 1, 2, \cdots$ and the events $S_i; i = 1, 2, \cdots$ are all ancillary.

In particular this is so if $\mathcal{E}^{S_i}$ is at least as informative as $\mathcal{E}; i = 1, 2, \cdots$.

Remark
One might of course also consider more general partitionings. Thus if Z is a statistic then we might for the various possible realizations z of Z consider the conditional experiments $\mathcal{E}^z$ given $Z = z$. Under general conditions which are described in Torgersen (1976) it can't occure that these conditional experiments are all at least as informative as $\mathcal{E}$ for the Hellinger ordering (and thus for the over all ordering) unless Z is ancillary and $\mathcal{E}^Z$ is equivalent to $\mathcal{E}$ with probability 1.

Proof:
The Hellinger transform of $\mathcal{E}^{S_i}$ is for a probability distribution $\lambda \in \Lambda$ given by: $H(\lambda|\mathcal{E}^{S_i}) = \int_{S_i} (\prod_\theta dP_\theta^{\lambda_\theta}) / \prod_\theta P_\theta(S_i)^{\lambda_\theta}$. The inequality $H(\lambda|\mathcal{E}^{S_i}) \leqq H(\lambda|\mathcal{E})$ may thus be written

$$\int_{S_i} \prod_\theta dP_\theta^{\lambda_\theta} \leqq \prod_\theta P_\theta(S_i)^{\lambda_\theta} H(\lambda|\mathcal{E}).$$

Assuming this for $i = 1, 2, \cdots$ we find by summation that

$$H(\lambda|\mathcal{E}) \leqq \sum_i \prod_\theta P_\theta(Si)^{\lambda_\theta} H(\lambda|\mathcal{E}) \leqq H(\lambda|\mathcal{E})$$

so that "=" prevails. By regularity $H(\lambda|\mathcal{E}) > 0$ that $\sum_i \prod_\theta P_\theta(S_i)^{\lambda_\theta} = 1$. Thus $P_\theta(S_i)$ does not depend on θ as long as $\lambda_\theta > 0$. Assuming that $H(\lambda|\mathcal{E}^{S_i}) \leqq H(\lambda|\mathcal{E})$ for all λ and

Then if $I(\theta,\mathcal{E}^{S_j}) \geqq I(\theta,\mathcal{E})$ for all θ and j then $I(\theta,\mathcal{E}^{S_j}) = I(\theta,\mathcal{E}); j = 1,2,\cdots$ and the events $S_1, S_2, \cdots$ are all ancillary.

Let us consider the particular case of an exponential experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ having θ as a k-dimensional canonical parameter. Thus we assume that there is a non negative measure μ and a k-dimensional statistic Y such that, for each $\theta, dP_\theta/d\mu = c(\theta)^{-1}e^{(\theta,Y)}h$ where the function h is non negative. The natural parameter set $\tilde{\Theta}$ of $\mathcal{E}$ is the set of vectors θ such that $\int e^{(\theta,Y)}hd\mu < \infty$. By Hölder's inequality $\tilde{\Theta}$ is convex and thus contains the convex hull of Θ. We shall consider the quantity $c(\theta) = \int e^{(\theta,Y)}hd\mu$ as defined for all $\theta \in \tilde{\Theta}$.

Let us agree to use the notation θ_λ for the expectation of θ for the prior distribution λ.

The Hellinger transform of $\mathcal{E}$ becomes for a prior probability distribution λ in Λ:

$$H(\lambda|\mathcal{E}) = \int \prod_\theta dP_\theta^{\lambda_\theta} = \int e^{(\theta_\lambda,Y)}hd\mu / \prod_\theta c(\theta)^{\lambda_\theta}$$
$$= c(\theta_\lambda)/\prod_\theta c(\theta)^{\lambda_\theta}$$

so that

$$\log H(\lambda|\mathcal{E}) = \log c(\theta_\lambda) - \sum_\theta \lambda_\theta \log c(\theta)$$

showing directly the log convexity of the function c.

If $P_\theta(S) > 0$ for some θ then $P_\theta(S) > 0$ for all θ and the selection experiment $\mathcal{E}^S$ is also exponential. In fact

$$dP_\theta(\cdot|S)/d\mu = \frac{1}{c(\theta)}P_\theta(S)^{-1}e^{(\theta,Y)}hI_S.$$

Hence

$$H(\lambda|\mathcal{E}^S) = H(\lambda|\mathcal{E})P_{\theta_\lambda}(S)/\prod_\theta P_\theta(S)^{\lambda_\theta}.$$

Thus $\mathcal{E}^S$ majorizes $\mathcal{E}$ or is majorized by $\mathcal{E}$ for the Hellinger ordering according to whether $\log P_\theta(S)$ is convex or is concave in θ.

As shown by Bayarri and DeGroot (1987) this is also the criterion for the ordering by Fisher information. In fact if θ is an interior point of Θ then the Fisher information matrix at θ is the convariance matrix of Y and the (i,j)th element of this matrix at θ is the number $\partial^2 \log c(\theta)/\partial\theta_i\partial\theta_j$.

Adding the matrix whose (i,j)th element is $\partial^2 \log P_\theta(S)/\partial\theta_i\partial\theta_j$ we obtain the Fisher information matrix of $\mathcal{E}^S$. Thus $I(\theta,\mathcal{E}^S) \geqq I(\theta,\mathcal{E})$ or $\leqq I(\theta,\mathcal{E})$ as $\log P_\theta(S)$ is locally convex or concave at θ.

(iii) $\log (E_\theta w_2 / E_\theta w_1)$ is convex in θ.

If Θ is open then these conditions are equivalent with:

(iv) $$I(\cdot, \mathcal{E}_{w_1}) \leqq I(\cdot, \mathcal{E}_{w_2}).$$

Remark
By Janssen's convolution criterion, theorem 1.1, $\mathcal{E}_{w_1} \leqq \mathcal{E}_{w_2}$ if and only if for some θ the distribution P_{θ,w_1} is a convolution factor of the distribution P_{θ,w_2}.

Let us return to the case where Θ is one dimensional and let us assume that Θ is an open interval of the real line and that $\mathcal{E}$ is exponential as above. By section 5 the experiment $\mathcal{E}$ is locally at least (at most) as informative as $\mathcal{E}^S$ if and only if $\mathcal{E}$ is pairwise at least (at most) as informative as $\mathcal{E}^S$. By the results of section 8 this is so if and only if $\mathcal{E}$ is at least (at most) as informative $\mathcal{E}^S$ for monotone decision problems and this, in turn, implies the corresponding orderings of Hellinger transforms and of Fisher information.

Having established one way or another an ordering of the experiment $\mathcal{E}$ and the selection experiment $\mathcal{E}^S$ we may ask whether or not this is the overall ordering of these experiments.

Consider e.g. the case of the Poisson distribution with the zero class missing. Thus $P_\theta(x) = \frac{\lambda^x}{x!} e^{-\lambda}; x = 0, 1, 2 \cdots$ with $\lambda = e^\theta; \theta \in R$ and $S = \{1, 2 \cdots\}$ so that $P_\theta(S) = 1 - e^{-\lambda}$. Then $P_\theta(S)$ is concave in θ so that, as argued in Bayarri and DeGroot (1987), selection on S decreases information pairwise and thus, by section 8, also for monotone decision problems.

Bayarri and DeGroot show also that $\mathcal{E}$ is not more informative than $\mathcal{E}^S$ in this case. We may here argue this on the basis of theorem 1.1. Indeed if $\mathcal{E}$ was more informative than $\mathcal{E}^S$ then $P_\theta(\cdot|S)$ must be a convolution factor of P_θ. P_θ being a Poisson distribution does not have other convolution factors than translates of Poisson distributions. As $P_\theta(\cdot|S)$ is clearly not a translate of a Poisson distribution we are forced to conclude that $\mathcal{E}$ is not more informative than $\mathcal{E}^S$.

Furthermore, since

$$dP_1(\cdot|S)/dP_0(\cdot|S) = [dP_1/dP_0][P_0(S)/P_1(S)]$$

is monotonically increasing in T:

$$\beta^S(\alpha) = P_1(T > c|S) + \gamma P_1(T = c|S)$$

so that

$$P_1(S)\beta^S(\alpha) = P_1(T \in]c, b)) + \gamma P_1(T = c) = [E_1\delta - P_1(T \in (b, \infty])$$
$$= \beta(E_0\delta) - P_1(T \in (b,\infty]) = \beta((1-\alpha)z_0 + \alpha z_1) - \beta(z_0) \text{ or } = \beta(\alpha z_1)$$

as

$$\infty \notin (a, b) \text{ or } \infty \in (a, b).$$

Altogether this shows that

$$\beta^S(\alpha) \equiv_\alpha [\beta((1-\alpha)z_0 + \alpha z_1) - \beta(z_0)]/[\beta(z_1) - \beta(z_0)] \text{ when } \infty \notin (a, b)$$

while

$$\beta^S(\alpha) \equiv_\alpha \beta(\alpha z_1)/\beta(z_1) \text{ when } \infty \in (a, b).$$

The latter expression is clearly also valid when $z_0 = 0$ and $\beta(0) = 0$. Note that the above formulas are precisely those we obtain from the representation (uniform (0,1), β) by using, respectively, $]z_0, z_1[$ and $[0, z_1]$ as the selection set.

Consider in particular selection sets of the form $S = [T \in (a, \infty]]$. If we put $z = P_0(S) = P_0(T \in (a, \infty])$ then, by the last one of the formulas above,

$$\beta^S(\alpha) \equiv_\alpha \beta(\alpha z)/\beta(z).$$

As the N-P function β^S of $\mathcal{D}^S$ depends on z only we shall permit ourselves to write β_z instead of β^S so that $\beta_z(\alpha) \equiv_\alpha \beta(\alpha z)/\beta(z)$. Conditions ensuring monotonicity in z of β_z provide, since $\beta_1 = \beta$, also conditions ensuring the informational inequalities $\mathcal{E}^S \geqq \mathcal{E}$ and $\mathcal{E}^S \leqq \mathcal{E}$.

As a particular case let us consider the exponential life time model realized by observing T having cumulative distribution function $F_\theta(t) = 1 - e^{\theta t}; t > 0$ where $\theta < 0$. Assuming that $\theta_1 \leqq \theta_2$ and putting $p(\theta_1, \theta_2) = |\theta_2|/|\theta_1|$ we find that $\beta(\alpha|\theta_1, \theta_2) \equiv_\alpha \alpha^{p(\theta_1,\theta_2)}$. If the selection set is the set $S = [T \geqq a]$ then, by the lack of "memory" of the exponential distribution, $P_\theta(\cdot|S) = P_\theta$ and thus $\mathcal{E}^S \sim \mathcal{E}$. Indeed $\beta_z(\alpha) \equiv_\alpha \beta(\alpha)$ whenever β is of the form $\beta(\alpha) \equiv_\alpha \alpha^p$ with $0 \leqq p \leqq 1$. In view of proposition 6.6 we may ask whether it is at all possible to have $\mathcal{D}^S \geqq \mathcal{D}$ for $\mathcal{D}^S$ not equivalent to $\mathcal{D}$.

If $K(x) = 1$ for some $x < \infty$ then $\int (dP_1/dP_0)^n dP_0 \leqq x^n < \infty; n = 1,2,\cdots$. Hence, by the proof of proposition 6.6, either $P_1 \perp P_0$ or $z = P_0(S) \geqq P_1(S) = \beta(z)$ for all $z > 0$. In the first case K is the one point distribution in 0. In the second case $\beta(\alpha) \equiv \alpha$ and then K is the one point distribution in 1. Assuming that K is not a one point distribution we conclude that the cumulative distribution function K is strictly increasing on I. If, furthermore, $0 < c_1 < c_2 \in I$ then $K(c_1-) < K(c_2-) < 1$. Putting $z_2 = 1 - K(c_1-)$ and $z_1 = 1 - K(c_2-)$ we find that $0 < z_1 < z_2$ and hence $\omega(z_1) \leqq \omega(z_2)$ i.e:

$$E[(X/c_2)|X \geqq c_2] = \int_{[c_2,\infty[} xK(dx)/c_2K[c_2,\infty[= \beta(z_1)/[K^{-1}(1-z_1)z_1]$$

$$= \omega(z_1)^{-1} \geqq \omega(z_2)^{-1} = E[(X/c_1)|X \geqq c_1].$$

Thus the equivalent conditions (a) and (b) imply condition (c).

Assume finally that condition (c) is satisfied. If, in addition, K is the one point distribution in a point ξ then $\beta(\alpha) \equiv \alpha\xi + (1-\xi)$ and then $\beta_z(\alpha) = [\alpha z\xi + (1-\xi)]/[z\xi + (1-\xi)]$ decreases monotonically in z. (Alternatively we may observe that $\omega(z) = \xi z/\beta(z) \uparrow$ in z). Thus we may, and shall, in the remaining part of the proof assume that K is not a one point distribution. Decompose the interval $]0,1[$ as $]0,1[= \bigcup\{[^*K(c-),K(c)] : c \geqq K^{-1}(0+)\}$ where the $*$ indicates that this bracket shall be reversed when $c = K^{-1}(0+)$. Letting J_c denote the closed interval with endpoints $P(X > c)$ and $P(X \geqq c)$ this decomposition may be expressed as $]0,1[= \bigcup\{J_c^* : c \geqq K^{-1}(0+)\}$ where the $(*)$ indicates that $P(X \geqq c)$ shall be deleted from J_c when $c = K^{-1}(0+)$.

If $c_1 < c_2$ then J_{c_1} and J_{c_2} are disjoint and J_{c_1} is entirely to the right of J_{c_2}. If $z \in J_c$ then $K^{-1}(1-z) = c$ and thus $\omega(z) = cz/\beta(z)$ which is monotonically increasing in z.

Assume so that $0 < z_1 < z_2 < 1$. Then $z_1 \in J_{c_2}$ and $z_2 \in J_{c_1}$ for numbers c_1 and c_2. By the above result we may assume that $c_1 \neq c_2$ and then $c_1 < c_2$. We may now also assume that $z_1 = 1 - K(c_2-)$ and that $z_2 = 1 - K(c_1)$. Then $K^{-1}(1-z_1) = c_2, K^{-1}(1-z_2) = c_1, \beta(z_1) = \int_{[c_2,\infty[} xK(dx)$ and $\beta(z_2) = \int_{]c_1,\infty[} xK(dx)$.

Letting $d \downarrow c_1$ we find that:

$$\omega(z_1)^{-1} = E[(X/c_2)|X \geqq c_2] \geqq \lim E[(X/d)|X \geqq d] = E[(X/c_1)|X > c_1] = \omega(z_2)^{-1}$$

so that $\omega(z_1) \leqq \omega(z_2)$.

□

Thus the property of $\mathcal{D}$ that $\mathcal{D}^S$ is information decreasing in $S = [dP_1/dP_0 \geqq a]$ may be phrased as $\mathcal{L}_K(\log X)$ having increasing mean exponential residual life time. Similarly, as we now shall see, the property of $\mathcal{D}$ that $\mathcal{D}^S$ is information increasing in S may be interpreted as $\mathcal{L}_K(\log X)$ having decreasing mean exponential residual life time.

$c_{1,n} = K^{-1}(p_1 + \frac{1}{n})$ when $p_1 + \frac{1}{n} < p_2$. Then $c_{1,n} \downarrow c_1$. It follows, by continuity, that it suffices to show that $E(X/c_{1,n}|X \geqq c_{1,n}) \geqq E(X/c_2|X \geqq c_2)$. Put $z_1 = 1 - K(c_2)$ and $z_{2,n} = 1 - K(c_{1,n})$. Then $0 < z_1 \leqq z_{2,n} < 1$ and $K^{-1}(1 - z_1) = c_2$ while $K^{-1}(1 - z_{2,n}) = c_{1,n}$. Thus

$$E(X/c_{1,n}|X \geqq c_{1,n}) = \frac{\beta(z_{2,n})}{z_{2,n}K^{-1}(1 - z_{2,n})} \geqq \frac{\beta(z_1)}{z_1 K^{-1}(1 - z_1)}$$
$$= E(X/c_1|x \geqq c_1).$$

Assume finally that (c) is satisfied and let us write $\omega(z) = zK^{-1}(1 - z)/\beta(z)$. Then $\omega(z) \leqq 1$ for all $z \in]0, 1[$.

As conditions (a) and (b) are trivially satisfied when K is either the one point distribution in 0 or the one point distribution in 1 we may, and shall, assume that K is none of these one point distributions.

Let $z \in]0, 1[$ and put $c_z = K^{-1}(1 - z)$. Then $c_z \in [K^{-1}(0+), K^{-1}(1)]$. Decompose $]0, 1[$ as $]0, 1[= U \cup V \cup W$ where $U = \{z : c_z = K^{-1}(1)\}, V = \{z : K^{-1}(0+) < c_z < K^{-1}(1)\}$ and $W = \{z : c_z = K^{-1}(0+)\}$. Put $b = K^{-1}(1)$.

If $z \in U = \{z : c_z = b\}$ then $K^{-1}(1 - z) = b$ and $1 = K(b) > 1 - z \geqq K(b-)$ so that $z \leqq K([b, \infty[) = K(\{b\})$. It follows that we may write $z = \theta K(\{b\})$ with $\theta \in]0, 1]$. Thus $\beta(z) = \theta b K(\{b\})$ so that $\omega(z) = 1$.

If $z \in V = \{z : K^{-1}(0+) < c_z < b\}$ then, since $c_z \in I$, we find that

$$\omega(z) = zK^{-1}(1 - z)/\beta(z) = [1 - K(c_z)]c_z/\beta(1 - K(c_z))$$
$$= 1/E[(X/c_z)|X \geqq c_z] \downarrow \text{ in } z.$$

Thus ω is monotonically decreasing on V.

If $z \in W = \{z : c_z = K^{-1}(0+)\} =]1 - K(K^{-1}(0+), 1[$ then $K(K^{-1}(0+)) > 0$ so that K has an atom at $K^{-1}(0+)$. By condition (c) this requires that either $K^{-1}(0+) = 0$ or that $K^{-1}(0+) = K^{-1}(1)$.

The last condition implies that K is a one point distribution and thus, since $\int xK(dx) = 1$, that K is the one point distribution in 1. Having excluded this case we conclude that $K^{-1}(0+) = 0$ and thus that $\omega(z) = c_z = 0$ when $z \in W$.

The proof is now completed by observing that U is situated entirely to the left of V while W is situated entirely to the right of V.

□

Let $r_\theta(x) = P_\theta(x)/P_\theta([x,\infty[)$ be the hazard rate at x. Then P_θ has the *IFR* property, i.e. $\mathcal{L}_\theta(X - m|X \geqq m)$ decreases monotonically in $m \in I$ according to the "stochastic ordering" of distribution functions, if and only if $r_\theta(x)$ increases monotonically in $x \in I$. In other words $P_\theta(X \geqq m + t|X \geqq m)$ decreases in m for each $t \geqq 0$ if and only if the hazard rate is monotonically decreasing on I. This can only happen when I is bounded from below.

A simple condition ensuring this is that $v(x+1)/v(x)$ is decreasing in $x \in I$. Note that if these conditions are satisfied then they are also satisfied after conditioning w.r.t any non empty right tail $[a,\infty[\cap I$ of I. The last condition prevails however after conditioning w.r.t any sub interval of I and also by translations and reflections. It may be checked that this last condition is satisfied in the (negative) binomial case as well as in the Poisson case.

Theorem 7.4 (Pairwise information increasing selections. The discrete case).

Let $\mathcal{E} = (P_\theta : \theta \in \Theta)$ be an exponential family of IFR distributions as described above.

Consider selection on an subinterval S of I (necessarily bounded from below). Then $\mathcal{E}^S$ is pairwise information increasing in S provided we restrict attention to intervals S having the same right end point as I. This provision may be omitted if $v(x+1)/v(x)$ decreases monotonically in x when $x \in I$.

Proof:
Let X be a random variable having distribution P_θ when θ prevails.

We may without loss of generality assume that I is a finite interval having the origin as its left and point. Passage to a general bounded interval follows then from the remarks above. Passage to unbounded intervals follows from Scheffe's convergence theorem and the fact that orderings of experiments are preserved under limits for the weak experiment topology. We may also restrict attention to $\theta \in \{\theta_0, \theta_1\}$ where θ_0 and θ_1 are two given points in Θ which we may assume are arranged so that $\phi(\theta_0) \leqq \phi(\theta_1)$. Simplifying the notations we shall write: $\phi_0 = \phi(\theta_0), \phi_1 = \phi(\theta_1), k_0 = k(\theta_0), k_1 = k(\theta_1), P_0 = P_{\theta_0}$ and $P_1 = P_{\theta_1}$. After conditioning we obtain the distributions Q_0 and Q_1 on $\{1,2,\cdots\}$ where

$$Q_i(x) = P_i(x)/1 - P_i(0); i = 0,1.$$

It suffices then to show that

$$\|(P_1 - cP_0)^+\| \geqq (Q_1 - cQ_0)^+\|$$

for any number c. If $c \leqq 0$ this is trivial. It follows from the convexity of these two functions of c that we may restrict attention to numbers c belonging to a given support of $K = \mathcal{L}_{P_0}(dP_1/dP_0)$. Thus we may assume that $c = P_1(a)/P_0(a)$ where $a \in I$. Putting $\Omega_P(a) = \|(P_1 - \frac{P_1(a)}{P_0(a)}P_0)^+\|$ and $\Omega_Q(a) = \|(Q_1 - \frac{P_1(a)}{P_0(a)}Q_0)^+\|$ the desired inequality may be

Assuming that $a+1 \in I$ the left hand side may be written:

$$E(\frac{P_1(X)}{P_0(X)}|X \geqq a+1) = E(\frac{P_1(X-a-1+a+1)}{P_0(X-a-1+a+1)}|X \geqq a+1)$$

and this, by the IFR property is

$$\leqq E_0[\frac{P_1(X-1+a+1)}{P_0(X-1+a+1)}|X \geqq 1] = E_0[\frac{P_1(X+a)}{P_0(X+a)}|X \geqq 1]$$
$$= \frac{P_0(0)}{P_1(0)}\frac{P_1(a)}{P_0(a)}\frac{1-P_1(a)}{1-P_0(0)}$$

proving our claim.

□

In the other direction we have:

Theorem 7.5 (Pairwise information decreasing selections. The discrete case).

Let $\mathcal{E} = (P_\theta : \theta \in \Theta)$ be an exponential family of DFR distributions on an interval I of the integers. (Then I is necessarily unbounded from above unless it is a one point set).

We shall assume that there are functions k and ϕ on Θ and a positive function v on I such that

$$P_\theta(x) = k(\theta)v(x)e^{\phi(\theta)x}; x \in I.$$

Then selection on an interval $S = [a, \infty[; -\infty \leqq a < \infty$ is pairwise information increasing in a; i.e pairwise information decreasing in S.

Remark:
P_θ has the DFR property if and only if $\mathcal{L}_\theta(X-m|X \geqq m)$ increases stochastically in $m \in I$ and this is equivalent to the condition that the hazard rate $P_\theta(x)/P_\theta[x,\infty[$ decreases monotonically in x.

A sufficient condition for P_θ to have the DFR property is that $v(x+1)/v(x)$ increases monotonically in $x \in I$.

Proof:
We may again restrict attention to the case where $I = [0,\infty[, S = [1,\infty[$ and $\Theta = \{\theta_0, \theta_1\}$ with $\phi(\theta_0) \leqq \phi(\theta_1)$. Using the notation of the proof of the previous theorem we put

$$k_0 = k(\theta_0), k_1 = k(\theta_1), \phi_0 = \phi(\theta_0), \phi_1 = \phi(\theta_1), P_0 = P_{\theta_0}, P_1 = P_{\theta_1}$$

and

$$Q_i(x) = P_i(x)/1 - P_i(0); i = 0, 1, x = 1, 2, \cdots.$$

It follows that $\Omega^*_P(a) \leqq \Omega^*_Q(a)$ if and only if

$$P_1([a+1,\infty[)/P_0([a+1,\infty[) \geqq [P_0(0)/P_1(0)][P_1(a)/P_0(a)].$$

The left hand side of the last inequality may also be expressed as:

$$\begin{aligned}&E_0[P_1(X)/P_0(X)|X \geqq a+1]\\&= E_0[P_1(X-a-1+a+1)/P_0(X-a-1+a+1)|X \geqq a+1].\end{aligned}$$

By the DFR property the last quantity is at least

$$\begin{aligned}&E_0[P_1(X+a+1)/P_0(X+a+1)]\\&= [P_1(a)/P_0(a)](E_0(k_1/k_0)e^{(\phi_1-\phi_0)X})(k_0/k_1)e^{\phi_1-\phi_0}\\&= [P_1(a)/P_0(a)][E_0(P_1(X)/P_0(X))][P_0(0)/P_1(0)]e^{\phi_1-\phi_0}\\&\geqq [P_1(a)/P_0(a)][1][P_0(0)/P_1(0)][1] \text{ since } \phi_1 \geqq \phi_0.\end{aligned}$$

□

Let us return to the general problem of the effect of selection for a M-L experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ with $\Theta \subseteqq R$. If our concern are with comparison for monotone decision problems then the preceding results may be helpful provided we know how to handle the dichotomies $(P_{\theta_1}, P_{\theta_2}) : \theta_1, \theta_2 \in \Theta$. Even if this is possible however it may be simpler to consider the local effect of selection and then apply the results of section 5.

Consider an experiment $\mathcal{E} = (P_\theta : \theta \in \Theta)$ which is differentiable in a given point $\theta \in \Theta$. Let us denote the slope function $\kappa(\cdot|\theta, \mathcal{E})$ by $\kappa(\cdot|\theta)$.

The selection experiment $\mathcal{E}^S = (P_\theta(\cdot|S) : \theta \in \Theta)$ is also differentiable in θ and

$$P^\bullet_\theta(A|S) = P^\bullet_\theta(A)/P_\theta(S) - P_\theta(A|S) \cdot [P^\bullet_\theta(S)/P_\theta(S)]$$

when $A \subseteqq S$. Thus

$$dP^\bullet_\theta(\cdot|S)/dP_\theta(\cdot|S) = dP^\bullet_\theta/dP_\theta - P^\bullet_\theta(S)/P_\theta(S)$$

Let $\kappa^S(\cdot|\theta)$ denote the slope function of $\mathcal{E}^S$ at θ.

Assume that $dP^\bullet_\theta/dP_\theta$ is a monotonically increasing function of some real valued statistic T and that the selection set S is of the form $S = [T \in (a,b)]$ where (a,b) denotes a specific interval with endpoints a and b. Express S as $S = [T \in (a,\infty)] - [T \in (b,\infty)]$ with the appropriate assignments of endpoints of intervals. Putting $z_0 = P_\theta(T \in (b,\infty))$ and $z_1 = P_\theta(T \in (a,\infty))$ we find that $P_\theta(S) = z_1 - z_0$ while $P^\bullet_\theta(S) = \kappa(z_1|\theta) - \kappa(z_0|\theta)$.

Proposition 7.6 (The particular role of doubly exponential translation experiments).

With notations as above we can't have $\kappa^S(\cdot|\theta) \geqq \kappa(\cdot|\theta)$ for $0 < z_0 < z_1 < 1$ unless $\kappa(z_0|\theta) = \kappa(z_1|\theta)$ and $\mathcal{L}(dP_\theta^\bullet/dP_\theta|P_\theta)$ either is a two point distribution or is the one point distribution in zero. If so then $\kappa^S(\cdot|\theta) = \kappa(\cdot|\theta)$ i.e. $\mathcal{E}^S$ and $\mathcal{E}$ are locally equivalent at θ.

Remark:
By example 3.1 the density $dP_{\theta_0}^\bullet/dP_{\theta_0}$ assumes P_{θ_0}- essentially at most two values if and only if the experiment $\mathcal{E}$ is locally, at $\theta = \theta_0$ equivalent with a doubly exponential translation family $(G(\cdot - \theta) : \theta \in R)$ with density g given by $g(x) = [(AB)/(A+B)]e^{-Bx^+ + Ax^-}; x \in R$. In that case $\kappa(\alpha) \equiv B\alpha \wedge A(1-\alpha)$. If selection is on the set $S = [a, b]$ and $\theta_0 = 0$ then $\kappa(z_0|0) = \kappa(z_1|0)$ if and only if $Aa + Bb = 0$. In particular $a = -b$ when $A = B$. Thus selection on any symmetric interval $[-b, b]$ preserves local information at zero for a symmetric double exponential translation family.

Proof:
The inequality $\kappa^S \geqq \kappa$ may, when $0 < \alpha \leqq 1$, be written

$$\kappa(\alpha)/\alpha \leqq \{[\kappa(z_\alpha) - \kappa(z_0)]/(z_\alpha - z_0)\} - [\kappa(z_1) - \kappa(z_0)]/(z_1 - z_0)$$

where $z_\alpha = (1-\alpha)z_0 + \alpha z_1$. By concavity $\kappa(\alpha)/\alpha \geqq [\kappa(z_\alpha) - \kappa(z_0)]/(z_\alpha - z_0)$ when $\alpha \leqq z_0$ and then $[\kappa(z_1) - \kappa(z_0)]/(z_1 - z_0) \leqq 0$ i.e. $\kappa(z_1) \leqq \kappa(z_0)$. Put $\kappa^*(z) = \kappa(1-z); 0 \leqq z \leqq 1$.

Replacing κ, z_0, z_1 and α with respectively $\kappa^*, 1-z_1, 1-z_0$ and $1-\alpha$ the same argument shows that $\kappa^*(1-z_0) \leqq \kappa^*(1-z_1)$ i.e. that $\kappa(z_0) \leqq \kappa(z_1)$. Hence $\kappa(z_0) = \kappa(z_1)$ and we denote this number by m.

The first inequality of this proof shows that for all α :

$$\frac{\kappa(\alpha)}{\alpha} \leqq \frac{\kappa(z_\alpha) - \kappa(z_0)}{z_\alpha - z_0}$$

so that "=" holds for $\alpha \leqq z_0$. If so then, by concavity, $\frac{\kappa(z_\alpha)-\kappa(z_0)}{z_\alpha - z_0} \leqq \frac{\kappa(z_0)}{z_0}$ so that $\frac{\kappa(\alpha)}{\alpha} \leqq \frac{\kappa(z_0)}{z_0}$ when $z \leqq z_0$. Thus $\kappa(\alpha) = \frac{\kappa(z_0)}{z_0}\alpha = \frac{m}{z_0}\alpha$ when $0 \leqq \alpha \leqq z_0$. By similar arguments, or by symmetry, $\kappa(\alpha) = \frac{m}{1-z_1}(1-\alpha)$ when $z_1 \leqq \alpha \leqq 1$.

It remains to investigate the behaviour of κ on $[z_0, z_1]$. Put $z_* = \frac{1}{1-z_1}/[\frac{1}{z_0} + \frac{1}{1-z_1}] = \frac{z_0}{1-(z_1-z_0)}$. Then $z_* \in]z_0, z_1[$ and $(1-z_*)/(1-z_1) = z_*/z_0$. By concavity $[\kappa((1-\alpha)z_0 + \alpha z_1) - m]/(z_1 - z_0) \geqq \kappa(\alpha)$ for all $\alpha \in [0,1]$. If $\alpha = z_*$ then $(1-\alpha)z_0 + \alpha z_1 = z_*$ yielding

$$\frac{\kappa(z_*) - m}{z_1 - z_0} \geqq \kappa(z_*) \text{ i.e.:}$$

$$\kappa(z_*)/z_* \leqq \frac{\kappa(z_*) - m}{z_* - z_0}$$

These conditions are automatically satisfied when F has the decreasing failure rate property DFR i.e. log $(1 - F)$ is convex.

Proof:
Let X be a random variable having distribution F and put $\omega(z) = \frac{\kappa(z)}{z} - F^{-1}(1-z); 0 < z < 1$. Then $\omega(z) \geqq 0$ for all $z \in]0,1[$.

Assume that conditions (i) and (ii) are fulfilled. The identity $\kappa(z)/z \equiv F^{-1}(1-z) + \omega(z)$ and the continuity of κ implies that

$$\kappa(z)/z = F^{-1}((1-z)+) + \omega(z-).$$

By monotonicity $F^{-1}(1-z) \leqq F^{-1}(1-z)+)$ and $\omega(z) \leqq \omega(z-)$ so that $=$ prevails. Thus F^{-1} and ω are both continuous on $]0,1[$. If $F^{-1}(0+) < c_1 < c_2 < F^{-1}(1)$ and $F(c_1) = F(c_2) = p$ then $0 < p < 1$. Putting $a = \inf\{x : F(x) = p\}$ we find that $F^{-1}(p) = a$ while $F^{-1}(q) \geqq c_2 > a$ whenever $q > p$. This, however, contradicts the continuity of F^{-1}. Thus F is strictly increasing on $]F^{-1}(0+), F^{-1}(1)[$.

It suffices therefor, in order to show that F is strictly increasing on I, to show that $F(x) < 1$ for all numbers x unless F is a one point distribution. Assume then that F is not a one point distribution and that this was not so. Then $a = F^{-1}(1) < \infty$. Put $z_\epsilon = F[a-\epsilon, \infty[= 1 - F(a-\epsilon)$ when $a - \epsilon$ is a point of continuity of F. Then $z_\epsilon \downarrow F[a, \infty[$ as $\epsilon \downarrow 0$. Consider first the situation where F assigns a positive mass $F(\{a\})$ to a. Then $F(\{a\}) = F[a, \infty[> 0$. It follows that if $z \in]0, F(\{a\})[$ then we may write $z = \theta F(\{a\}) + F]a, \infty[$ so that $\kappa(z) = \theta a F(\{a\}) + \int_{]a,\infty[} xF(dx)$. The rightmost terms of the last two equalitites are both zero. Hence $\kappa(z)/z = a$. Furthermore $F^{-1}(1-z) = F^{-1}(F(a-)) = a$ and thus $\omega(z) = a - a = 0$. In particular $\omega(0+) = 0$. Suppose next that $a = F^{-1}(1) < \infty$ and that a is a point of continuity of F. Then, as $\epsilon \downarrow 0, z_\epsilon \downarrow F[a, \infty[= 0$ and $\omega(z_\epsilon) \downarrow (0+)$. But

$$\omega(z_\epsilon) = E(X|X \geqq a - \epsilon) - F^{-1}(1 - z_\epsilon) = E(X|X \geqq a - \epsilon) - (a - \epsilon) \leqq a - (a - \epsilon) = \epsilon \to 0.$$

Thus we find again that $\omega(0+) = 0$. Hence, since ω is monotonically decreasing, $\omega(z) \equiv 0$. It follows that

$$0 \equiv \kappa(z) - zF^{-1}(1-z) \equiv \int_0^z [F^{-1}(1-t) - F^{-1}(1-z)]dt$$

so that F^{-1} is constant on $]0,1[$. As $\int xF(dx) = 0$ this constant must be zero contradicting our assumption that F is not the one point distribution in zero.

Altogether this shows that F is strictly increasing on I when conditions (i) and (ii) are fulfilled. Still assuming this put $z_c = 1 - F(c-)$ when $c \in I$. Then z_c is strictly increasing in $c \in I$ and $E(X - c|X \geqq c) = [\kappa(z_c)/z_c] - F^{-1}(1 - z_c) = \omega(z_c) \uparrow$ in c. Thus conditions (i) and (ii) imply condition (iii).

$m = P_r(X = \xi)$. Put also $z(\theta) = \theta m + 1 - F(\xi)$ for $\theta \in]0,1[$. Then, by the Neyman-Pearson lemma,

$$\kappa(z(\theta)) = \theta\xi m + \int_{]\xi,\infty[} xF(dx) = \theta\xi m + \kappa(1 - F(\xi)).$$

If $m > 0$ then $z(\theta) \in]0,1[$ for all $\theta \in]0,1[$ and:

$$\kappa(z(\theta))/z(\theta) = \omega(z(\theta)) + F^{-1}(1 - z(\theta)) = \omega(z(\theta)) + \xi.$$

By (ii) $\omega(z(\theta))$ increases monotonically with θ. On the other hand $z \to \kappa(z)/z$ is always monotonically decreasing. Thus $\kappa(z(\theta))/z(\theta)$ does not depend on $\theta \in]0,1[$. Letting $\theta_1 \to 0$ and $\theta_2 \to 1$ in the identity $\kappa(z(\theta_1))z(\theta_2) = \kappa(z(\theta_2))z(\theta_1)$ we find that $\kappa(z(0))[m + z(0)] = [\xi m + \kappa(z(0))]z(0)$ so that: $\kappa(z(0))/z(0) = \xi$. It follows that $0 \leqq \kappa(z)/z = \xi$ when $z \in [1 - F(\xi), 1 - F(\xi-)]$ and thus, by concavity, $\kappa(z) = \xi z$ whenever $0 < z \leqq 1 - F(\xi-)$.

It follows that $\xi = \kappa'(0+)$. Thus, since $\kappa'(z) \leqq \kappa'(0+) = \xi$ for all $z, P_r(X \leqq \xi) = 1$. Hence, since $P_r(X = \xi) > 0, F^{-1}(1) = \xi$ which is contrary to our assumption that $\xi < F^{-1}(1)$. Thus F is nonatomic on J.

Decompose J as

$$J =]-\infty, F^{-1}(0+)[\cup \bigcup_{0<z<1} [F^{-1}(1-z), F^{-1}((1-z)+)].$$

If $-\infty < c \leqq F^{-1}(0+)$ then $E(X - c|X \geqq c) = EX - c = -c \downarrow$ in c and we obtain the smallest value $-F^{-1}(0+)$ when $c = F^{-1}(0+)$. If $c \in J_z =]F^{-1}(1-z), F^{-1}((1-z)+)[$ then $F(c) = 1 - z$. It follows that $E(X|X \geqq c)$ does not depend on c as long as $c \in J_z$.

Thus $E(X - c|X \geqq c) = E(X|X \geqq c) - c$ is decreasing in $c \in \overline{J}_z = [F^{-1}(1-z), F^{-1}((1-z)+)]$ whenever $0 < z < 1$. Put $c_z = F^{-1}(1-z)$ when $0 < z < 1$. Then $E(X - c_z|X \geqq c_z) = \omega(1 - F(c_z))$ which is monotonically increasing in z.

Consider so numbers $c_1 < c_2$ in the interval $]F^{-1}(0+), F^{-1}(1)[$. If

$$F(c_1) = F(c_2) = 1 - z \text{ then } c_1, c_2 \in [F^{-1}(1-z), F^{-1}((1-z)+)]$$

and $E(X - c_1|X \geqq c_1) \geqq E(X - c_2|X \geqq c_2)$. On the other hand if $1 - z_1 = F(c_1) < F(c_2) = 1 - z_2$ and if $\overline{c}_1 = F^{-1}((1-z_1)+)$ and if $z \uparrow z_1$ then:

$$\begin{aligned} E(X - c_1|X \geqq c_1) &\geqq E(X - \overline{c}_1|X \geqq \overline{c}_1) = \lim_z E(X - c_z|X \geqq c_z) \\ &\geqq E(X - c_{z_2}|X \geqq c_{z_2}) \geqq E(X - c_2|X \geqq c_2). \end{aligned}$$

Finally, noting that $E(X - c|X \geqq c) \to -F^{-1}(0+)$ as $c \downarrow F^{-1}(0+)$, we conclude that $E(X - c|X \geqq c)$ is monotonically decreasing in $c \in J$.

Note that the above arguments remain valid if our assumptions are relaxed by replacing the factor $e^{\theta x}$ in the density by $e^{\phi(\theta)x}$ where ϕ is differentiable with $\phi^\bullet(\theta) \neq 0$.

In particular the arguments apply with θ being replaced by $-\theta$ and for the family of distribution $\mathcal{L}(-X|P_\theta) : \theta \in \Theta$.

It follows in all these cases that selection on an interval S is locally information increasing in S. Hence forth this selection is also information increasing for monotone decision problems.

One might be tempted to infer from the last example that if $\mathcal{E} = (G_\theta : \theta \in \Theta)$ is a differentiable strongly unimodal translation family then selection on an interval S is locally information increasing in S. Expressing the continuous density g of G as $g = e^{-\omega}$ where ω is continuous and convex on R we see that $-g'/g = \omega'$ is monotonically increasing. Now $-g'(x)/g(x) \in S$ if and only if x belongs to the interval $\{y : \omega'(y) \in S\}$. Furthermore if $\mathcal{E}$ is realized by observing X and if I is an interval then the events $[X \in I]$ and $[-g'(X)/g(X) \in$ the interval spanned by $\omega'[I]]$ are equivalent. The local behaviour of $\mathcal{E}$ is determined by $F = \mathcal{L}(-g'(X)/g(X)|G)$ and, as explained in section 3, F may prescribe any kind of local behaviour which is not associated with the zero slope function. This carries over to the local effect of selection on intervals as well. Thus selection on intervals may in this case be locally information decreasing, increasing or neither.

In the discrete case the last theorems are not applicable. Let us however again consider the situation described in the last statement of theorem 7.4. Thus $\mathcal{E} = (P_\theta : \theta \in \Theta)$ is an exponential model with sample space I being an interval of integers and P_θ given by:

$$P_\theta(x) = k(\theta)v(x)e^{\phi(\theta)x}; x \in I$$

where ϕ is some function on Θ and v is a positive function on I such that $v(x+1)/v(x) \downarrow$ in $x \in I$. We may without loss of generality assume that $\phi(\theta) \equiv \theta$ and we shall assume that Θ is open.

In order to show that $\mathcal{E}^S$ is information increasing in the sub interval S of I the crucial case is, as explained before theorem 7.4, the case where I has the origin as its left end point and the selection set is $S = I - \{0\}$.

We shall now see how this case may be simply argued by "local" comparison. Let X be a random variable whose distribution θ is P_θ when θ prevails. Thus $X \geqq 0$ and selection is to the set $S = [X \geqq 1]$. Then $d/d\theta \log P_\theta(x) = x - E_\theta X$ so that

$$F = \mathcal{L}(X - E_\theta X).$$

In the selection experiment we arrive at

$$\tilde{F} = \mathcal{L}(X - (E_\theta X|X \geqq 1)|X \geqq 1).$$

In order to show that F is a dilation of $\tilde{F}$ we must show that

$$\int (x-c)^{\pm} F(dx) \geqq \int (x-c)^{\pm} \tilde{F}(dx); c \in R$$

8. APPROXIMATE COMPARISON.

In his pathbreaking 1964 work on sufficiency and approximate sufficiency LeCam introduced a notion of approximate comparison. For experiments $\mathcal{E}$ and $\mathcal{F}$ and a non negative function ϵ on Θ this notion is as follows:

Consider some decision problem specified by a decision space T and a loss function L. Consider also a risk function s obtainable in $\mathcal{F}$ for this decision problem. If we in such a situation always (and thus for any (T,L)) may ensure that there is a risk function r obtainable in $\mathcal{E}$ which dominates s (i.e. $r \leqq s$) then $\mathcal{E}$ is at least as informative as $\mathcal{F}$. Indeed this is one of a series of several possible equivalent definitions of "being at least as informative as".

According to the terminology established in section 6 an experiment $\mathcal{E}$ is at least as informative as an experiment $\mathcal{F}$ if and only if for every loss function L the decision problem $(\mathcal{E}, L)$ is at least as informative as the decision problem $(\mathcal{F}, L)$. If so then we write $\mathcal{E} \geqq \mathcal{F}$ so that $\mathcal{E} \geqq \mathcal{F}$ if and only if $(\mathcal{E}, L) \geqq (\mathcal{F}, L)$ for every loss function L.

If there is no risk function r in $\mathcal{E}$ which dominates s then one might hope for a risk function r in $\mathcal{E}$ such that $r(\theta) \leqq s(\theta) + \epsilon(\theta)\sup_t |L_\theta(t)|; \theta \in \Theta$. Here $L_\theta(t)$ is the loss incurred by the decision t when θ prevails and the loss function L applies. The number $\sup_t |L_\theta(t)|$ is a normalizing quantity expressing the "size" of the loss function at θ. When convenient this quantity may, as we here shall, be replaced with other expressions for this "size".

If $\epsilon(\theta) \geqq 2$ the above inequality holds trivially for this particular θ. In general one might hope to find small numbers $\epsilon(\theta) : \theta \in \Theta$ such that there to any decision problem (it suffices to consider finite decision problems) and to any risk function s obtainable in $\mathcal{F}$ there is a risk function r in $\mathcal{E}$ such that:

$$r(\theta) \leqq s(\theta) + \epsilon(\theta)\sup_t |L_\theta(t)|; \theta \in \Theta.$$

If this is so, for a given function ϵ, then according to the definition in LeCam (take or give a factor 2) the experiment $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$.

His basic randomization criterion states that $\mathcal{E} = (P_\theta : \theta \in \Theta)$ is ϵ-deficient w.r.t. $\mathcal{F} = (Q_\theta : \theta \in \Theta)$ if and only if $\|P_\theta T - Q_\theta\| \leqq \epsilon_\theta; \theta \in \Theta$ for some transition (= generalized Markov kernel) from $\mathcal{E}$ to $\mathcal{F}$.

By the terminology used in section 6 the experiment $\mathcal{E}$ is ϵ-deficient w.r.t. the experiment $\mathcal{F}$ if and only if for every loss function L the decision problem $(\mathcal{E}, L)$ is κ_ϵ deficient w.r.t. the decision problem $(\mathcal{F}, L)$ where $\kappa_\epsilon(\theta) = \sup_t |L_\theta(t)|\epsilon(\theta); \theta \in \Theta$.

If we restrict attention to a particular collection τ of loss functions then we obtain a concept of ϵ-deficiency of $\mathcal{E}$ w.r.t. $\mathcal{F}$ for τ. The smallest constant $\epsilon = \delta_{(\tau)}(\mathcal{E}, \mathcal{F})$ such that $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ (for τ) is *the deficiency of* $\mathcal{E}$ *w.r.t* $\mathcal{F}$ *(for* τ*)*. The largest of

(iv) $(P_{\theta_0}, P_{\theta_1})$ is $(\epsilon(\theta_0), \epsilon(\theta_1))$ deficient w.r.t. $(Q_{\theta_0}, Q_{\theta_1})$ for any pair (θ_0, θ_1) of points from Θ.

If these conditions are satisfied then the decision procedure in $\mathcal{E}$ producing the risk function r in condition (i) may be chosen independently of the monotone loss function L and thus only depending on the decision procedure in $\mathcal{F}$ producing the risk function s.

Remark 1:
(iii) is just a reformulation of (ii). The powerfunction π in $\mathcal{E}$ may always be chosen within the set $\prod_{\mathcal{E}}$ of powerfunctions of most powerful tests.

Note also that we in (iv) may restrict attention to pairs (θ_0, θ_1) such that $\theta_0 < \theta_1$.

Remark 2:
The pseudonorm $\| \quad \|_*$ is related to the supremum norm $\| \quad \|$ by the inequalities:

$$\|L_\theta\|_* \leqq 2\|L_\theta\|$$

and

$$\sup_{t_1,t_2} |L_\theta(t_2) - L_\theta(t_1)| \leqq 2\|L_\theta\|_*$$

It folllows that condition (i) implies that $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ for monotone decision problems with non negative loss functions. On the other hand if this is so then (i) holds with ϵ_θ replaced throughout with $2\epsilon_\theta$.

Proof:
By the above remarks (iii) $\Longleftrightarrow$ (ii) $\Longrightarrow$ (iv) and trivially (i) $\Longrightarrow$ (ii). Hence (i) $\Longrightarrow$ (ii) $\Longleftrightarrow$ (iii) $\Longrightarrow$ (iv). The proof will be completed by first establishing the implication (iv) $\Longrightarrow$ (iii) and then showing that (iii) $\Longrightarrow$ (i).

We may, see the introduction, assume that the parameter set Θ is finite.

Assume so that condition (iv) is satisfied. If $\#\Theta \leqq 2$ then (iii) is just a reformulation of (iv). Suppose then that we have argued the implication $(iv) \Rightarrow (iii)$ when $\#\Theta = n$ and let us consider the situation where $\#\Theta = n + 1$, say $\Theta = \{0, 1, \cdots, n\}$. Let $\theta_0 = j$ where $0 \leqq j \leqq n$ and consider the problem of testing "$\theta \leqq \theta_0$" against "$\theta > \theta_0$" for the 0-1 loss function.

Consider first the case "$j = 0$". Deleting $\theta = n - 1$ from Θ we conclude, from the induction hypothesis, that there is a powerfunction $\tilde{\rho}$ in $\prod_{\mathcal{E}}$ (defined in section 4) so that $\tilde{\rho}(0) \leqq \sigma(0) + \frac{1}{2}\epsilon_0$ while $\tilde{\rho}(i) \geqq \sigma(i) - \frac{1}{2}\epsilon_i; i = 1, \cdots, n-2, n$. If $\tilde{\rho}(n-1) \geqq \sigma(n-1) - \frac{1}{2}\epsilon_{n-1}$

Remembering that $\prod_{\mathcal{E}}$ is totally ordered we may construct $\rho_1, \cdots, \rho_{k-1}$ so that $\rho_1 \geqq \rho_2 \geqq \cdots \geqq \rho_{k-1}$.

We may achieve this by replacing $\rho_i, i = 1, \cdots, k-1$, with $\tilde{\rho}_i = \rho_i \vee \cdots \vee \rho_{k-1}$. If $\theta \leqq \theta_i$ and $k-1 \geqq j \geqq i$ then $\theta \leqq \theta_j$ and $\rho_j(\theta) \leqq E_\theta\delta_j + \epsilon_\theta/2 \leqq E_\theta\delta_i + \epsilon_\theta/2$. Hence $\tilde{\rho}_i(\theta) \leqq E_\theta\delta_i + \epsilon_\theta/2$ when $\theta \leqq \theta_i$. If $\theta > \theta_i$ then $\tilde{\rho}_i(\theta) \geqq \rho_i(\theta) \geqq E_\theta\delta_i + \epsilon_\theta/2$.

Let us therefor assume that $\rho_0 \geqq \rho_1 \geqq \rho_2 \geqq \cdots \geqq \rho_{k-1} \geqq \rho_k$ where $\rho_0 = 1$ and $\rho_k = 0$. Then there are testfunctions $1 = \phi_0 \geqq \phi_1 \geqq \cdots \geqq \phi_{k-1} \geqq \phi_k = 0$ such that $E_\theta\phi_i \equiv_\theta \rho_i(\theta)$. Put finally $\psi_i = \phi_{i-1} - \phi_i; i = 1, \cdots, k$. Then $\psi_i, \cdots, \psi_k \geqq 0$ and $\psi_1 + \cdots + \psi_k = 1$. The test functions $\psi_1, \cdots, \psi_k$ define the decision procedure ψ in $\mathcal{E} = (P_\theta : \theta \in \Theta)$ given by: $\psi(t_i|\cdot) = \psi_i; i = 1, \cdots, k$. The risk function r of ψ may be expressed as:

$$r(\theta) \equiv_\theta \sum_{i=0}^{k-1}[L_\theta(t_{i+1}) - L_\theta(t_i)]E_\theta\phi_i.$$

Hence

$$r(\theta) - s(\theta) = \sum_{i=1}^{k-1}[L_\theta(t_{i+1}) - L_\theta(t_i)](E_\theta\phi_i - E_\theta\delta_i).$$

(The 0-th term may be disregarded since $E_\theta\phi_0 = E_\theta\delta_0 = 1$).

Assume so that $\theta_{j-1} < \theta \leqq \theta_j$. If $i < j$ then $L_\theta(t_i) \geqq L_\theta(t_{i+1})$ and, since $\theta > \theta_i, E_\theta\phi_i \geqq E_\theta\delta_i - \frac{1}{2}\epsilon_\theta$. Hence

$$\begin{aligned}[L_\theta(t_{i+1}) - L_\theta(t_i)](E_\theta\phi_i - E_\theta\delta_i) &= [L_\theta(t_i) - L_\theta(t_{i+1})](E_\theta\delta_i - E_\theta\phi_i) \\ &\leqq \frac{1}{2}\epsilon_\theta[L_\theta(t_i) - L_\theta(t_{i+1})] \text{ when } i < j.\end{aligned}$$

If $i \geqq j$ then $L_\theta(t_{i+1}) \geqq L_\theta(t_i)$ and, since $\theta \leqq \theta_i, E_\theta\phi_i \leqq E_\theta\delta + \frac{1}{2}\epsilon_\theta$. Hence

$$[L_\theta(t_{i+1}) - L_\theta(t_i)](E_\theta\phi_i - E_\theta\delta_i) \leqq \frac{1}{2}\epsilon_\theta[L_\theta(t_{i+1}) - L_\theta(t_i)]$$

when $i \geqq j$. It follows that

$$\begin{aligned} r(\theta) - s(\theta) &= \sum_{i=1}^{k-1} = \sum_{i=1}^{j-1} + \sum_{i=j}^{k-1} \\ &\leqq \frac{1}{2}\epsilon_\theta[(L_\theta(t_1) - L_\theta(t_j)) + ((L_\theta(t_k) - L_\theta(t_j))] \leqq \epsilon_\theta\|L_\theta\|_*.\end{aligned}$$

A similar analysis shows that the inequality $r(\theta) - s(\theta) \leqq \epsilon_\theta\|L_\theta\|_*$ is also valid when $\theta \leqq \theta_1$ or $\theta > \theta_{k-1}$.

□

REFERENCES.

Arnold, B. 1986. Majorization and the Lorenz Order: A brief introduction. Lecture notes in statistics, 43, Springer-Verlag.

Bayarri, M.J. and DeGroot, M.H. 1987. Information in selection models. In probability and Bayesian statistics. (R. Viertl, ed.), 39-51, New York: Plenum.

Bayarri, M.J. and DeGroot, M.H. 1989. Comparison of experiments with weighted distributions. Technical report No. 450. Dept. of statistics, Carnegie Mellon University, Pittsburgh.

Blackwell, D. 1951. Comparison of experiments. Proc. Second Berkeley Sympos. Math. Statist. Probab. 93-102.

Blackwell, D. 1953. Equivalent comparisons of experiments. Ann. Math. Statist. 24, 265-272.

C. Boll, 1955. Comparison of experiments in the infinite case. Ph.D. thesis. Stanford University.

Ehm, W. and Müller, D.W. 1983. Factorizing the information contained in an experiment conditionally on the observed value of a statistic. Z. Wahrscheinlichkeitstheorie verw. Geb. 65, 121-134.

Janssen, A. 1988. A convolution theorem for the comparison of exponential families. Res. report no 211, Univ. of Siegen.

Karlin, S. and Rubin, H. 1956. The theory of decision procedures for distributions with monotone likelihood ratio. Ann. Math. Statist 27, 272-299.

Klefsjö, B. 1984. Reliability interpretations of some concepts from economics. Naval Research Logistic Quarterly 31, 301-308.

LeCam, L. 1964. Sufficiency and approximate sufficiency. Ann. Math. Statist. 35, 1419-1455.

LeCam, L. 1986. Asymptotic methods in statistical decision theory. Springer-Verlag.

Lehmann, E. 1988. Comparing location experiments. Ann. Statist. 16, 521-533.

Loéve, M. 1977. Probability theory I, II, Springer-Verlag.

Marshall, A.W. and Olkin, I. 1979. Inequalities: Theory of majorization and its applications. Academic Press.

Menger, K. 1952. The formative years of Abraham Wald and his work in geometry. Ann. Math. Statist. 23, 14-20.